



TESIS - TE142599

**KLASIFIKASI DATA KEUANGAN SEKTOR PUBLIK  
UNTUK PENENTUAN SAMPEL PEMERIKSAAN  
MENGUNAKAN *K-NEAREST NEIGHBORS***

AHMAD DWI ARIANTO  
NRP 2215206715

DOSEN PEMBIMBING  
Dr. Ir. Achmad Affandi, DEA  
Dr. Supeno Mardi Susiki Nugroho, S.T., M.T.

PROGRAM MAGISTER  
BIDANG KEAHLIAN TELEMATIKA - CIO  
DEPARTEMEN TEKNIK ELEKTRO  
FAKULTAS TEKNOLOGI ELEKTRO  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2017





TESIS - TE142599

**KLASIFIKASI DATA KEUANGAN SEKTOR PUBLIK  
UNTUK PENENTUAN SAMPEL PEMERIKSAAN  
MENGUNAKAN *K-NEAREST NEIGHBORS***

AHMAD DWI ARIANTO  
NRP 2215206715

DOSEN PEMBIMBING  
Dr. Ir. Achmad Affandi, DEA  
Dr. Supeno Mardi Susiki Nugroho, S.T., M.T.

PROGRAM MAGISTER  
BIDANG KEAHLIAN TELEMATIKA - CIO  
DEPARTEMEN TEKNIK ELEKTRO  
FAKULTAS TEKNOLOGI ELEKTRO  
INSTITUT TEKNOLOGI SEPULUH NOPEMBER  
SURABAYA  
2017



## LEMBAR PENGESAHAN

Tesis disusun untuk memenuhi salah satu syarat memperoleh gelar  
Magister Teknik (M.T)  
di  
Institut Teknologi Sepuluh Nopember

oleh:

Ahmad Dwi Arianto  
NRP. 2215206715

Tanggal Ujian : 6 Juni 2017  
Periode Wisuda: September 2017

Disetujui oleh:

1. Dr. Ir. Achmad Affandi, DEA (Pembimbing I)  
NIP: 196510141990021001
2. Dr. Supeno Mardj S.N., S.T., M.T. (Pembimbing II)  
NIP: 197003131995121001
3. Dr. Ir. Wirawan, DEA (Penguji)  
NIP: 196311091989031011
4. Dr. I Ketut Eddy Purnama, S.T., M.T. (Penguji)  
NIP: 196907301995121001
5. Dr. Adhi Dharma Wibawa, S.T., M.T. (Penguji)  
NIP: 197605052008121003



Dekan Fakultas Teknologi Elektro

Dr. Tri Arief Sardjono, S.T., M.T.  
NIP. 197002121995121001

*Halaman ini sengaja dikosongkan*

## PERNYATAAN KEASLIAN TESIS

Dengan ini saya menyatakan bahwa isi keseluruhan Tesis saya dengan judul “**KLASIFIKASI DATA KEUANGAN SEKTOR PUBLIK UNTUK PENENTUAN SAMPEL PEMERIKSAAN MENGGUNAKAN *K-NEAREST NEIGHBORS***” adalah benar-benar hasil karya intelektual mandiri, diselesaikan tanpa menggunakan bahan-bahan yang tidak diijinkan dan bukan merupakan karya pihak lain yang saya akui sebagai karya sendiri.

Semua referensi yang dikutip maupun dirujuk telah ditulis secara lengkap pada daftar pustaka. Apabila ternyata pernyataan ini tidak benar, saya bersedia menerima sanksi sesuai peraturan yang berlaku.

Surabaya, Mei 2017



Ahmad Dwi Arianto

NRP 2215206715

*Halaman ini sengaja dikosongkan*



# **KLASIFIKASI DATA KEUANGAN SEKTOR PUBLIK UNTUK PENENTUAN SAMPEL PEMERIKSAAN MENGUNAKAN *K-NEAREST NEIGHBORS***

Nama mahasiswa : Ahmad Dwi Arianto  
NRP : 2215206715  
Pembimbing : 1. Dr. Ir. Achmad Affandi, DEA  
2. Dr. Supeno Mardi Susiki Nugroho, S.T., M.T.

## **ABSTRAK**

Penentuan sampel pemeriksaan internal yang dilakukan oleh Inspektorat Utama BPK RI seringkali berdasarkan “kebiasaan” semata. Metode penambangan data bisa menjadi salah satu cara alternatif untuk mewariskan “kebiasaan” tersebut pada generasi penerus. Penelitian deteksi opini dan deteksi pemalsuan/kecurangan laporan keuangan menggunakan teknik penambangan data (*data mining*) sudah sangat lazim di sektor swasta, namun tidak di sektor publik. Orientasi pada laba menjadi pembeda utama yang bermuara pada karakter laporan keuangan kedua sektor tersebut. Klasifikasi dilakukan melalui dua tahap, tahap pertama klasifikasi berdasarkan 2 kelas, dan tahap kedua klasifikasi berdasarkan 4 kelas. Validasi silang 10 lipatan (*10-folds cross validation*) yang dipadukan dengan berbagai skema normalisasi data, fitur, jarak, Nilai-K, dan uji lanjut akan digunakan untuk membangun model terbaik. Hasil penelitian menunjukkan bahwa area di bawah kurva/*Area Under Curve (AUC)* tertinggi untuk klasifikasi KNN 2 kelas adalah 66,30%, sedangkan AUC untuk 4 kelas adalah 61,56% yang termasuk klasifikasi dengan kinerja buruk. Penelitian lebih lanjut diperlukan untuk meningkatkan kinerja model prediksi opini dari penelitian ini.

Kata kunci: sampel, penambangan data, opini, laporan keuangan sektor publik, KNN

*Halaman ini sengaja dikosongkan*

# **CLASSIFICATION OF PUBLIC SECTOR FINANCIAL DATA FOR DETERMINING THE AUDIT SAMPLES USING K-NEAREST NEIGHBORS**

By : Ahmad Dwi Arianto  
Student Identity Number : 2215206715  
Supervisor(s) : 1. Dr. Ir. Achmad Affandi, DEA  
2. Dr. Supeno Mardi S.N., S.T., M.T.

## **ABSTRACT**

The determination of the internal audit sample conducted by the Main Inspectorate of BPK RI is often based on the mere “habits”. Data mining can be one of the alternative ways to pass “the habits” on the next generation. Research on opinion detection and falsified/fraudulent financial statement detection using data mining techniques is very common in the private sector, but not in the public sector. Profit orientation was a key differentiator that led to the different characteristics of financial statements in the both sector. The classification was done through two stages, firstly classification based on 2 classes, and secondly classification based on 4 classes. Ten folds cross-validation combined with various data normalization schemes, features schemes, distances schemes, K-values schemes, and posthoc test schemes were used to build the best model. The results showed that The highest Area Under Curve (AUC) for classification KNN 2 class was 66.30%, while the highest AUC for 4 classes was 61.56% which fell under poor classification. Further research is needed to improve the performance of classification models from this study.

Keywords: sample, data mining, opinion, public sector financial statement, KNN

*Halaman ini sengaja dikosongkan*

## KATA PENGANTAR

Alhamdulillah, atas berkat rahmat Allah jua-lah tesis ini selesai. Salawat dan salam bagi Nabi Muhammad SAW Sang Pembimbing Umat. Terima kasih kepada Bapak, Ibu, Istri, Anak, Adik dan semua pihak, baik yang disebut atau tidak dalam tesis ini, yang telah memberikan dukungan baik moril maupun materiil demi terselesaikannya tesis ini, *Jazakumullah Khoiron Katsiron*.

Penulis mengucapkan terima kasih kepada Bapak Dr. Ir. Achmad Affandi, DEA selaku pembimbing pertama dan Bapak Dr. Supeno Mardi S.N., S.T., M.T. selaku pembimbing kedua yang telah meluangkan waktu dan memberikan masukan sehingga tesis ini dapat selesai.

Penulis juga menyampaikan terima kasih yang tak terhingga kepada:

1. Badan Pemeriksa Keuangan Republik Indonesia dan Kementerian Komunikasi dan Informasi yang telah memberikan kesempatan mendapatkan beasiswa Program Magister (S2) Telematika/*Chief Information Officer* pada Institut Teknologi Sepuluh Nopember Surabaya.
2. Prof. Ir. Joni Hermana, M.Sc.Es, Ph.D., selaku Rektor Institut Teknologi Sepuluh Nopember Surabaya.
3. Dr. Tri Arief Sardjono, S.T., M.T., selaku Dekan Fakultas Teknologi Elektro, Institut Teknologi Sepuluh Nopember Surabaya.
4. Dr. Ir. Wirawan, DEA, selaku Kepala Program Studi Pascasarjana Fakultas Teknologi Elektro.
5. Dr. Adhi Dharma Wibawa, ST, MT, selaku Koordinator Bidang Keahlian Telematika/ *Chief Information Officer (CIO)* sekaligus Dosen Wali Akademik CIO 2015 Jurusan Teknik Elektro, atas kesabaran, arahan, dan bimbingan kepada kami semua.
6. Seluruh Pengajar dan staf Program Studi Magister (S2) Departemen Teknik Elektro, Bidang Keahlian Telematika/*Chief Information Officer (CIO)*, atas jasa dan pengabdian dalam mendidik dan mendewasakan kami.
7. Keluargaku tercinta Bunda Nining Setyowati, Bunda Kuraesin, Bunda Marpi'ah, Ayahanda Sunarto, Ayahanda Ali Muchtar, Istriku Anisy Syahidah,

Ananda Annisa Istifiyanza Bihurin'in Arianto, Adinda Ahmad Khoironi Arianto, Adinda Aisyah Kresnaningtyas Aisyah Khoirunnisa' Rizkiyah Rohmah Salman Al-Kahfi, serta seluruh keluarga Boyolali, Cimahi, dan Blora atas semua doa, cinta, dan kasih sayang kalian jua tesis ini selesai.

8. Rekan-Rekan CIO 2015 (Mas Harun, Mbak Erlin, Mas Didin, Mas Arif, Mbak Navik, Mbak Asri, Mbak Indira, Mas Mamad, Mbak Indah, Mbak Intan, Mbak Lia, Mbak Erna, Mbak Fulah, dan Mas Adi) atas semua “hal” yang kita lalui bersama.
9. Rekan-Rekan Tim Futsal Pascasarjana Elektro dan ITS atas semua keceriaan di lapangan dan luar lapangan yang kalian berikan (Mas Angga, Mas Jemi, Mas Adlian, Mas Abby dkk yang tidak dapat kami sebut satu persatu, thank a lot bro)

Surabaya, Mei 2017

Penulis

## DAFTAR ISI

LEMBAR PENGESAHAN .....	iii
PERNYATAAN KEASLIAN TESIS .....	v
ABSTRAK .....	vii
<i>ABSTRACT</i> .....	ix
KATA PENGANTAR .....	xi
DAFTAR ISI.....	xiii
DAFTAR GAMBAR .....	xvii
DAFTAR TABEL.....	xix
BAB 1 PENDAHULUAN .....	1
1.1 Latar Belakang .....	1
1.2 Rumusan Masalah.....	5
1.3 Tujuan .....	5
1.4 Batasan Masalah .....	5
1.5 Kontribusi .....	6
1.6 Metodologi Penelitian.....	6
BAB 2 KAJIAN PUSTAKA.....	7
2.1 Kajian Penelitian Terkait .....	7
2.2 Pengertian Laporan Keuangan.....	8
2.3 Pengertian Opini Laporan Keuangan.....	8
2.4 Data Laporan Keuangan .....	13
2.5 Penambangan Data ( <i>Data Mining/Knowledge Discovery in Database</i> ). ..	14
2.6 Metode Tetangga Terdekat/ <i>K-Nearest Neighbors</i> .....	15
2.6.1 Konsep Kedekatan .....	16
2.6.2 Tahapan Algoritma KNN.....	17
2.7 Metode Evaluasi Klasifikator .....	18
2.7.1 <i>Hold-Out</i> .....	18
2.7.2 Validasi Silang 10-Lipatan.....	19
2.8 Uji Wilcoxon Mann Whitney.....	19
2.9 Uji Kruskal-Wallis .....	23
2.10 Uji Friedman .....	27

2.11	Uji Wilcoxon Peringkat Bertanda ( <i>Wilcoxon Sign Rank</i> ) .....	28
2.12	<i>F-Measure</i> .....	28
2.13	Area di Bawah Kurva <i>ROC/AUC</i> .....	29
BAB 3 METODOLOGI PENELITIAN .....		31
3.1	Alur Penelitian.....	31
3.2	Pemerolehan Data Mentah .....	32
3.3	Normalisasi Data .....	36
3.3.1	Pembersihan Data .....	36
3.3.2	Skema Normalisasi Data .....	36
3.4	Alokasi Data .....	38
3.5	Penggunaan Skema Fitur .....	39
3.6	Penggunaan Skema Nilai-K .....	41
3.7	Penggunaan Skema Jarak .....	42
3.8	Perbandingan Kinerja Skema .....	42
BAB 4 HASIL DAN PEMBAHASAN .....		45
4.1	Statistik Deskriptif Data Penelitian .....	45
4.2	Hasil Proses Normalisasi Data .....	46
4.3	Hasil Proses Alokasi Data .....	47
4.3.1	Alokasi Data untuk Klasifikasi Dua Kelas .....	47
4.3.2	Alokasi Data untuk Klasifikasi Empat Kelas .....	47
4.4	Hasil Statistik Uji Signifikansi untuk Seleksi Fitur.....	48
4.4.1	Seleksi Fitur untuk Klasifikasi Dua Kelas.....	48
4.4.2	Seleksi Fitur untuk Klasifikasi Empat Kelas.....	50
4.5	Implementasi Hasil Uji Statistik pada Skema Fitur .....	53
4.6	Sepuluh Model Terbaik Berdasarkan Rerata Harmonik <i>F-Measure</i> dan AUC (RH) .....	54
4.6.1	Sepuluh Model Terbaik Klasifikasi 2 Kelas.....	54
4.6.2	Sepuluh Model Terbaik Klasifikasi 4 Kelas.....	55
4.7	Hasil Uji Friedman pada Skema Normalisasi Data.....	56
4.7.1	Hasil Uji SND pada Klasifikasi 2 Kelas.....	56
4.7.2	Hasil Uji SND pada Klasifikasi 4 Kelas.....	57
4.8	Hasil Uji Friedman pada Skema Fitur .....	59
4.8.1	Uji Friedman pada Skema Fitur Klasifikasi 2 Kelas .....	59



4.8.2	Uji Friedman pada Skema Fitur Klasifikasi 4 Kelas.....	60
4.9	Hasil Uji Wilcoxon Peringkat Bertanda pada Skema Nilai-K.....	62
4.9.1	Uji WSR pada Skema Nilai-K Klasifikasi 2 Kelas.....	62
4.9.2	Uji WSR pada Skema Nilai-K Klasifikasi 4 Kelas.....	63
4.10	Hasil Uji Wilcoxon Peringkat Bertanda pada Skema Jarak.....	64
4.10.1	Uji WSR pada Skema Jarak Klasifikasi 2 Kelas.....	64
4.10.2	Uji WSR pada Skema Jarak Klasifikasi 4 Kelas.....	65
4.11	Hasil Uji WSR pada Skema Uji Lanjut Tukey <i>HSD</i> dan WMW .....	66
4.11.1	Uji WSR pada Skema Uji Lanjut Klasifikasi 2 Kelas.....	66
4.11.2	Uji WSR pada Skema Uji Lanjut Klasifikasi 4 Kelas.....	67
4.12	Interpretasi Kinerja Klasifikasi Menggunakan AUC.....	68
4.13	Skema Sistem Alarm ( <i>Early Warning System</i> ) Anomali Opini.....	71
4.13.1	Skema Sistem Alarm Klasifikasi 2 Kelas .....	71
4.13.2	Skema Sistem Alarm Klasifikasi 4 Kelas .....	73
BAB 5 PENUTUP .....		77
5.1	Kesimpulan .....	77
5.2	Saran .....	77
DAFTAR PUSTAKA .....		79
LAMPIRAN.....		83

*Halaman ini sengaja dikosongkan*

## DAFTAR GAMBAR

Gambar 2.1 Diagram alir tahapan algoritma KNN .....	18
Gambar 3.1 Diagram alur penelitian .....	31
Gambar 3.2 Skema pemilihan alokasi data terbaik .....	39
Gambar 4.1 Kurva ROC untuk Klasifikasi 2 Kelas .....	70
Gambar 4.2 Kurva ROC untuk Klasifikasi 4 Kelas .....	71
Gambar 4.3 Skema Sistem Alarm Klasifikasi 2 Kelas .....	73
Gambar 4.4 Skema Sistem Alarm Klasifikasi 4 Kelas .....	75

*Halaman ini sengaja dikosongkan*

## DAFTAR TABEL

Tabel 1.1 Jumlah LHP LK .....	3
Tabel 2.1 Penelitian Terkait .....	7
Tabel 2.2 Data Umur Peserta pada Tiap Kelas .....	21
Tabel 2.3 Skor T untuk Data yang Sama pada Uji Wilcoxon Mann Whitney .....	22
Tabel 2.4 Hasil Pemeringkatan Data Umur dan Keputusan atas H <sub>0</sub> .....	22
Tabel 2.5 Data Umur Peserta pada Tiap Kelas .....	24
Tabel 2.6 Skor T untuk Data yang Sama pada Uji Kruskal-Wallis .....	25
Tabel 2.7 Hasil Pemeringkatan Data Umur dan Keputusan atas H <sub>0</sub> .....	25
Tabel 2.8 Interpretasi AUC .....	30
Tabel 3.1 Rincian Data LHP LK .....	32
Tabel 3.2 Modifikasi Variabel Keuangan Sektor Swasta .....	33
Tabel 3.3 Skema Normalisasi Data .....	38
Tabel 3.4 Skema Fitur Penelitian .....	40
Tabel 3.5 Ilustrasi Pengujian Perbedaan Pengaruh Skema Jarak .....	43
Tabel 4.1 Statistik Deskriptif Data Penelitian .....	45
Tabel 4.2 Data yang Mengalami Proses Pembersihan Data .....	46
Tabel 4.3 Rerata Akurasi Proses Alokasi Data untuk Dua Kelas .....	47
Tabel 4.4 Rerata Akurasi Proses Alokasi Data untuk Empat Kelas .....	47
Tabel 4.5 Fitur Signifikan Sesuai Uji Wilcoxon Mann Whitney .....	48
Tabel 4.6 Nilai Log <sub>10</sub> dari Nilai-P Uji Wilcoxon Mann Whitney .....	49
Tabel 4.7 Fitur Signifikan Sesuai Uji KW dan Tukey <i>HSD</i> .....	50
Tabel 4.8 Nilai Log <sub>10</sub> dari Nilai-P Uji KW dan Tukey <i>HSD</i> .....	51
Tabel 4.9 Fitur Signifikan lebih dari satu uji Sesuai Uji KW-Tukey <i>HSD</i> .....	51
Tabel 4.10 Fitur Signifikan Sesuai Uji KW dan Uji WMW .....	52
Tabel 4.11 Nilai Log <sub>10</sub> dari Nilai-P Uji KW dan Uji WMW .....	52
Tabel 4.12 Fitur Signifikan lebih dari satu uji Sesuai Uji KW-WMW .....	53
Tabel 4.13 Skema Fitur Penelitian Lengkap Setelah Uji Statistik .....	53
Tabel 4.14 Sepuluh Model Terbaik Klasifikasi 2 Kelas .....	55
Tabel 4.15 Sepuluh Model Terbaik Klasifikasi 4 Kelas .....	55
Tabel 4.16 Rerata dan Varian Peringkat SND Klasifikasi 2 Kelas .....	56
Tabel 4.17 Hasil Uji Friedman pada SND Klasifikasi 2 Kelas .....	57
Tabel 4.18 Rerata dan Varian Peringkat SND Klasifikasi 4 Kelas .....	58
Tabel 4.19 Hasil Uji Friedman pada SND Klasifikasi 4 Kelas .....	58
Tabel 4.20 Rerata dan Varian Peringkat Skema Fitur (2 Kelas) .....	59
Tabel 4.21 Perbedaan Signifikan pada Skema Fitur Klasifikasi 2 Kelas .....	60
Tabel 4.22 Rerata dan Varian Peringkat Skema Fitur (4 Kelas) .....	61
Tabel 4.23 Perbedaan Signifikan pada Skema Fitur Klasifikasi 4 Kelas .....	61
Tabel 4.24 Rerata dan Varian Peringkat Skema Nilai-K (2 Kelas) .....	62
Tabel 4.25 Hasil Uji WSR pada Skema Nilai-K Klasifikasi 2 Kelas .....	63
Tabel 4.26 Rerata dan Varian Peringkat Skema Nilai-K (4 Kelas) .....	63

Tabel 4.27 Hasil Uji WSR pada Skema Nilai-K Klasifikasi 4 Kelas.....	64
Tabel 4.28 Rerata dan Varian Peringkat Skema Jarak (2 Kelas).....	64
Tabel 4.29 Hasil Uji WSR pada Skema Jarak Klasifikasi 2 Kelas .....	65
Tabel 4.30 Rerata dan Varian Peringkat Skema Jarak (4 Kelas).....	65
Tabel 4.31 Hasil Uji WSR pada Skema Jarak Klasifikasi 4 Kelas .....	65
Tabel 4.32 Rerata dan Varian Peringkat Skema Uji Lanjut (2 Kelas) .....	66
Tabel 4.33 Hasil Uji WSR pada Skema Uji Lanjut Klasifikasi 2 Kelas .....	67
Tabel 4.34 Rerata dan Varian Peringkat Skema Uji Lanjut (4 Kelas) .....	67
Tabel 4.35 Hasil Uji WSR pada Skema Uji Lanjut Klasifikasi 4 Kelas .....	68
Tabel 4.36 AUC dari Skema Pengujian Lanjutan untuk 4 Kelas .....	69
Tabel 4.37 Model Dengan AUC Tertinggi pada Klasifikasi 2 Kelas dan 4 Kelas	69
Tabel 4.38 <i>Confusion Matrix</i> dan AUC pada Klasifikasi 2 Kelas.....	69
Tabel 4.39 <i>Confusion Matrix</i> dan AUC pada Klasifikasi 4 Kelas.....	70
Tabel 4.40 Tabel Bantu Skema Sistem Alarm Klasifikasi 2 Kelas .....	72
Tabel 4.41 Tabel Bantu Skema Sistem Alarm Klasifikasi 4 Kelas .....	74
Tabel 5.1 Fitur Publik Termodifikasi Berkarakter Sektor Swasta.....	78

# **BAB 1**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Kecurangan atau penipuan (*fraud*) dan korupsi telah berkembang pesat di seluruh dunia [1], *Association of Certified Fraud Examiners (ACFE)* melaporkan bahwa terdapat 2.410 kasus kecurangan terkait pelaksanaan pekerjaan yang diselidiki antara Januari 2014 dan Oktober 2015 yang berlangsung di 114 negara yang berbeda [2]. Penipuan dan korupsi menyerang seluruh dunia mulai dari negara berkembang sampai negara maju, bahkan menyerang negara yang dari dulu sudah dikenal sebagai negara yang “bersih”. Mereka tidak hanya menyerang sektor swasta yang berorientasi pada laba/keuntungan, tapi juga sektor publik yang berorientasi pada layanan publik. Kondisi ini menjadi perhatian utama para legislator di sektor publik atau pembuat kebijakan publik di seluruh dunia. Korupsi telah merampok dana publik sehingga kebijakan publik tidak bisa dilaksanakan sesuai rencana karena kelangkaan dana. Korupsi membunuh kepercayaan publik terhadap institusi politik dan pemerintah, menyebabkan pengabaian peraturan dan undang-undang, mendistorsi penggunaan dana publik, mematikan persaingan yang sehat dalam penyediaan barang dan jasa publik, serta menghalangi orang miskin mendapatkan layanan dasar yang penting/layanan publik yang menguasai hajat hidup orang banyak [1].

Penipuan dan korupsi merupakan pelanggaran kejujuran yang merupakan salah satu kode etik akuntan [3] [4]. Fakta menunjukkan bahwa kecurangan dan korupsi tidak hanya dilakukan oleh pihak manajemen, namun terkadang juga melibatkan akuntan publik, termasuk auditor [5] [6], kondisi ini menunjukkan bahwa kejujuran dihargai, namun tidak dieksekusi dengan baik. Kasus internasional yang paling populer adalah kasus Enron pada tahun 2001 yang melibatkan Kantor Akuntan Publik (KAP) Arthur Andersen, salah satu dari lima KAP terbesar di dunia saat itu. Kasus ini menyebabkan kehancuran KAP Arthur Andersen, sehingga sampai saat ini hanya dikenal empat KAP besar di dunia (Ernst & Young, Deloitte & Touche, KPMG, dan PriceWaterhouseCoopers) dengan menyisakan tragedi yang

menimpa KAP Arthur Andersen sebagai sejarah kelam yang pernah mewarnai dunia audit. KAP Arthur Andersen melanggar kode etik akuntan dengan memanipulasi laporan keuangan Enron. Manipulasi tersebut dilakukan dengan cara tidak melaporkan jumlah hutang sebenarnya dari Enron. Selain memanipulasi laporan keuangan Enron, KAP Arthur Andersen juga menghancurkan dokumen pendukung mengenai kebangkrutan yang melanda Enron [7]. Kasus pelanggaran kode etik akuntan lainnya yang terjadi di Indonesia adalah kasus penyuapan yang diterima oleh oknum auditor Badan Pemeriksa Keuangan Republik Indonesia (BPK-RI). Oknum auditor tersebut menerima suap atau sogokan untuk mengubah temuan audit yang mengindikasikan terjadinya kerugian negara pada proyek pengadaan peralatan di Balai Latihan Kerja Kementerian Tenaga Kerja dan Transmigrasi. Suap dilakukan untuk mengubah temuan audit berindikasi kerugian negara menjadi temuan audit mengenai kesalahan prosedural semata [6]. Kasus lain yang pernah melanda oknum auditor BPK-RI adalah penyuapan terkait perubahan opini atas laporan keuangan atau yang dikenal dengan istilah jual beli opini [5]. Auditor memiliki tanggung jawab moral menjadi "penjaga gerbang" untuk melindungi investasi publik, namun sering didapati auditor yang gagal menghargai peran mereka sebagai "penjaga gerbang" tersebut [8]. Alih-alih menjaga kepercayaan publik, terkadang auditor malah mengkhianati kepercayaan publik yang dibebankan pada mereka. Kondisi ini seperti yang dimaksudkan oleh sebuah peribahasa kuno Indonesia "bagaikan pagar yang makan tanaman".

Identifikasi pelanggaran etika yang dilakukan oleh auditor sangat sulit dilakukan. Kondisi ini bisa terungkap paling sering melalui pengaduan (*whistleblowing*) [2]. Kondisi yang lebih parah, namun sering terjadi, pelanggaran kode etik tersebut tidak dapat diidentifikasi oleh lembaga audit tempat oknum auditor tersebut bekerja, namun baru diketahui saat ditangani langsung oleh aparat penegak hukum melalui operasi penangkapan yang menyebabkan kehancuran kehormatan lembaga audit yang menjadi tempat bekerja oknum auditor tersebut.

Permasalahan lain adalah pengabaian sumber informasi atau pengetahuan potensial dalam bentuk dokumen hasil pemeriksaan (baik *softcopy* atau *hardcopy*). Dokumen tersebut dibiarkan terbelengkalai tanpa pernah diolah lebih lanjut untuk mendapatkan pengetahuan tersembunyi di dalamnya, padahal pengetahuan para



ahli dapat dijumpai salah satunya pada hasil karya mereka. Para auditor senior adalah ahli di bidang auditing/pemeriksaan dan laporan hasil pemeriksaan mereka adalah buah karya yang merupakan kristalisasi pengetahuan dan pengalaman selama puluhan tahun yang sayang jika hanya dibiarkan menumpuk berdebu di gudang atau “berkarat” di ruang penyimpanan digital (*harddisk*).

Setiap tahun BPK-RI melakukan pemeriksaan atas laporan keuangan. sehingga setiap tahun akan selalu bertambah data Laporan Hasil Pemeriksaan atas Laporan Keuangan (LHP LK). Berdasarkan Ikhtisar Hasil Pemeriksaan BPK RI Semester I Tahun 2014 s.d. Semester I Tahun 2016 [9] [10] [11] [12] [13], diketahui terdapat 1.914 LHP LK sebagaimana tersaji pada Tabel 1.1.

Tabel 1.1 Jumlah LHP LK

<b>Tahun/Entitas</b>	<b>Jumlah LHP LK</b>
<b>2014</b>	<b>632</b>
- BUMN & Badan Lainnya	19
- Pemerintah Pusat	87
- Pemerintah Daerah	526
<b>2015</b>	<b>642</b>
- BUMN & Badan Lainnya	6
- Pemerintah Pusat	97
- Pemerintah Daerah	539
<b>2016</b>	<b>640</b>
- BUMN & Badan Lainnya	7
- Pemerintah Pusat	100
- Pemerintah Daerah	533
<b>Total 2014 s.d. 2016</b>	<b>1.914</b>

Namun, jumlah data yang terus bertambah tersebut hanya disimpan dalam bentuk *softcopy* (pdf) atau *hardcopy* LHP dari tahun ke tahun tanpa pernah dicoba untuk dianalisis lebih lanjut guna mendapatkan pengetahuan tersembunyi dari data tersebut yang berguna bagi kepentingan BPK-RI.

Kecerdasan buatan memiliki relasi dengan ranah akuntansi dan auditing lebih dari 30 tahun yang lalu. Ekspansi penggunaan kecerdasan buatan dalam ranah akuntansi dimulai sejak 1980, hal ini ditandai dengan munculnya berbagai penelitian mengenai penggunaan kecerdasan buatan dalam bidang akuntansi

termasuk audit. Pertama kali penggunaan kecerdasan buatan diusulkan, diteliti, dan dikembangkan dalam bidang audit/pemeriksaan, perpajakan, akuntansi manajemen, dan akuntansi keuangan [14].

Kecerdasan buatan menawarkan metode alternatif untuk mengatasi masalah pengabaian/penumpukkan dokumen LHP dan masalah kesulitan dalam penentuan sampel atas LHP yang diduga mengalami risiko kesalahan opini atau perdagangan/jual beli opini yang dilakukan oleh auditor. Metode kecerdasan buatan seperti klasifikasi/prediksi dapat digunakan untuk memprediksi opini dari suatu laporan keuangan. Jika metode tersebut telah teruji akurasi, maka Inspektorat Utama (Itama) selaku auditor internal BPK RI dapat memanfaatkan metode tersebut untuk mengecek risiko terjadinya kesalahan opini hasil pemeriksaan laporan keuangan [5] [6] [15] [16].

Penelitian deteksi opini laporan keuangan [17] [18] atau penelitian lain terkait akuntansi yang lebih berat, semisal deteksi kecurangan/pemalsuan laporan keuangan [19] [20] [21] [22] dan deteksi anomali transaksi keuangan [23], menggunakan teknik penambangan data (*data mining*) sudah sangat lazim di sektor swasta, namun tidak di sektor publik. Orientasi pada laba menjadi pembeda utama yang bermuara pada karakter laporan keuangan kedua sektor tersebut. Penelitian deteksi opini laporan keuangan antara lain menggunakan metode klasifikasi menggunakan *Probabilistic Neural Network/PNN* [17] dan *K-Nearest Neighbors/KNN* [18]. Penelitian [17] dilakukan menggunakan 3.333 data laporan keuangan sektor swasta dengan hasil bahwa kinerja PNN jauh melebihi kinerja *Artificial Neural Networks/ANN* dan *Logistic Regression*. Penelitian [18] menggunakan 5.276 data laporan keuangan perusahaan, baik perusahaan swasta maupun perusahaan publik (seperti Badan Usaha Milik Negara/BUMN dan Badan Usaha Milik Daerah/BUMD di Indonesia) dengan hasil bahwa KNN lebih efisien/memiliki rerata akurasi lebih baik daripada *Discriminant Analysis* dan *Logit Analysis*.

*K-Nearest Neighbors* (KNN) adalah metode yang berbasis *Nearest Neighbors* (NN), salah satu metode yang paling tua, simpel, dan terkenal namun memiliki kinerja yang bagus [24]. *Nearest Neighbors* juga termasuk satu dari sepuluh metode yang paling banyak digunakan dalam penambangan data/*data*

*mining* [25]. Kinerja KNN bahkan mampu melampaui kinerja metode lain yang lebih rumit [24] [26] [27]. Akan tetapi, pemilihan Nilai-K menjadi masalah utama yang berpengaruh terhadap akurasi klasifikasi. Penelitian ini menggunakan metode klasifikasi dengan KNN untuk deteksi anomali opini laporan keuangan sektor publik. Anomali opini suatu laporan yang berhasil dideteksi dapat menjadi tanda awal terjadinya risiko kesalahan opini atau jual beli opini. Tanda awal ini dapat menjadi pertimbangan utama untuk memilih (uji petik) laporan yang akan diperiksa lebih dalam.

## **1.2 Rumusan Masalah**

Berdasarkan uraian latar belakang di atas rumusan masalah yang diajukan adalah adanya kesulitan Inspektorat Utama dalam menentukan sampel laporan keuangan yang akan diperiksa lebih lanjut untuk menghindari risiko kesalahan opini, padahal sumber data potensial berupa dokumen LHP menumpuk di gudang atau sekedar disimpan dalam bentuk *softcopy*. Selain itu, adanya kelangkaan penelitian terkait penerapan kecerdasan buatan di bidang akuntansi dan audit keuangan sektor publik di Indonesia.

## **1.3 Tujuan**

Penelitian ini bertujuan menghasilkan klasifikasi data keuangan yang dapat dijadikan masukan bagi Inspektorat Utama dalam penentuan sampel pemeriksaan menggunakan metode *K-Nearest Neighbors*. Selain itu, penelitian ini juga bertujuan mengisi kesenjangan penelitian terkait penerapan kecerdasan buatan di bidang akuntansi dan audit keuangan sektor publik di Indonesia.

## **1.4 Batasan Masalah**

Penelitian ini dilakukan untuk menentukan sampel pemeriksaan atas Laporan Keuangan (LK) sektor publik pada tingkat Pemerintah Daerah (Pemerintah Provinsi, Kabupaten dan Kota di Indonesia). Tidak mencakup LK Pemerintah Pusat (Kementerian dan Lembaga) dan LK BUMN dan Badan lainnya.

## **1.5 Kontribusi**

Hasil penelitian ini diharapkan dapat menjadi masukan bagi Inspektorat Utama (Itama) selaku auditor internal BPK RI untuk memilih LHP LK yang perlu diperiksa lebih lanjut terkait risiko kesalahan opini agar kredibilitas BPK RI tetap terjaga.

## **1.6 Metodologi Penelitian**

Penelitian ini menggunakan metode KNN baik untuk klasifikasi 2 kelas (Baik/WTP dan Buruk/selain WTP) ataupun 4 kelas (WTP, WDP, TW, TMP). Berbagai kombinasi skema (SND, fitur, Nilai-K, dan jarak) digunakan untuk mencari skema terbaik. Uji statistik (Friedman dan Wilcoxon Peringkat Bertanda/WSR) akan digunakan untuk penentuan skema terbaik. Rincian metode dan langkah penelitian dipaparkan pada Bab 3.

## BAB 2

### KAJIAN PUSTAKA

#### 2.1 Kajian Penelitian Terkait

Penelitian penggunaan kecerdasan buatan/*artificial intelligence* dan *data mining* di ranah akuntansi sektor swasta dan di luar negeri sangatlah banyak, namun tidak demikian halnya dengan penelitian serupa di Indonesia. Masih sangat jarang dijumpai penelitian menggunakan kecerdasan buatan atau penambangan data di ranah akuntansi sektor publik di Indonesia. Sehingga penelitian ini diharapkan menjadi pionir bagi kemunculan penelitian serupa di Indonesia. Berikut ini disajikan beberapa penelitian terkait yang mendukung penelitian ini sebagaimana tersaji pada Tabel 2.1.

Tabel 2.1 Penelitian Terkait

No	Peneliti	Judul	Jumlah Sampel	Variabel	Metode	Validasi	Hasil
1	Baldwin, Amelia A.; Brown, Carol E.; Trinkle, Brad S. (2006)	<i>Opportunities For Artificial Intelligence Development In The Accounting Domain: The Case For Auditing</i>	-		review analitis	-	Ada potensi kenaikan penggunaan kecerdasan buatan dalam penelitian akuntansi
2	Gaganis, Chrysovalantis; Pasiouras, Fotios; Spathis, Charalambos; Zopounidis, Constantin (2007)	<i>A comparison of nearest neighbours, discriminant and logit models for auditing decisions</i>	5.276	- <b>Dependen</b> (opini) - <b>Independen</b> (ROA, Likuiditas, pertumbuhan aset, ekuitas/total aset, rating, jenis industri)	KNN, Analisis Diskriminan, Analisis Logit	Holdout	akurasi KNN unggul dibandingkan dua metode lainnya
3	Gaganis, Chrysovalantis; Pasiouras, Fotios; Doumpos, Michael (2007)	<i>Probabilistic neural networks for the identification of qualified audit opinions</i>	3.333	-y (opini) -log biaya audit -log remunerasi direktur -log jumlah pegawai	PNN, ANN, regresi logistik	Holdout	kinerja PNN melampaui 2 metode lainnya

Tabel 2.1 Penelitian Terkait (Lanjutan)

No	Peneliti	Judul	Jumlah Sampel	Variabel	Metode	Validasi	Hasil
				dan 24 variabel lainnya			
4	Ravisankar, P.; Ravi, V.; Rao, G. Raghava; Bose, I. (2011)	<i>Detection Of Financial Statement Fraud And Feature Selection Using Data Mining Techniques</i>	202	-utang -total aset - laba kotor - dan 33 variabel lain	MLFF, SVM, GP, GMDH, LR, PNN	10-folds validation	kinerja PNN melampaui metode-metode lainnya

## 2.2 Pengertian Laporan Keuangan

Laporan Keuangan merupakan laporan pertanggungjawaban pelaksanaan APBN/APBD yang sekurang-kurangnya terdiri dari laporan realisasi anggaran, neraca, laporan arus kas dan catatan atas laporan keuangan yang penyusunannya harus sesuai dengan standar akuntansi pemerintah. Pemerintah Pusat dan Pemerintah Daerah wajib menyampaikan laporan keuangan yang telah diperiksa BPK kepada DPR/selambat-lambatnya 6 (enam) bulan setelah berakhirnya tahun anggaran yang bersangkutan [28].

## 2.3 Pengertian Opini Laporan Keuangan

Menurut Buletin Teknis Pelaporan Hasil Pemeriksaan Atas Laporan Keuangan Pemerintah definisi opini adalah: *“pernyataan profesional sebagai kesimpulan pemeriksa mengenai kewajaran informasi yang disajikan dalam laporan keuangan”*. Adapun jenis opini yang dapat diberikan oleh para pemeriksa yaitu [29]:

- Wajar Tanpa Pengecualian (WTP)/*Unqualified Opinion*** memuat suatu pernyataan bahwa laporan keuangan menyajikan secara wajar, dalam semua hal yang material sesuai dengan SAP. Laporan dengan predikat WTP berarti bebas dari salah saji material. Sesuai dengan Standar Profesional Akuntan Publik (SPAP) yang diberlakukan dalam Standar Pemeriksaan Keuangan Negara (SPKN), BPK dapat memberikan opini Wajar Tanpa Pengecualian dengan

Paragraf Penjelas (WTP-DPP)/*Unqualified Opinion With Explanatory Paragraph* karena keadaan tertentu sehingga mengharuskan pemeriksa menambahkan suatu paragraf penjelasan dalam LHP sebagai modifikasi opini WTP;

- b. **Wajar Dengan Pengecualian (WDP)/*Qualified Opinion*** memuat suatu pernyataan bahwa laporan keuangan menyajikan secara wajar, dalam semua hal yang material sesuai dengan SAP, kecuali untuk dampak hal-hal yang berhubungan dengan yang dikecualikan;
- c. **Tidak Wajar (TW)/*Adversed Opinion*** memuat suatu pernyataan bahwa laporan keuangan tidak menyajikan secara wajar dalam semua hal yang material sesuai dengan SAP;
- d. **Pernyataan Menolak Memberikan Opini atau Tidak Menyatakan Pendapat (TMP)/*Disclaimer of Opinion*** menyatakan bahwa pemeriksa tidak menyatakan opini atas laporan keuangan. Opini TMP bagi sebagian akuntan dianggap bukanlah suatu opini, karena pemeriksa menolak memberikan pendapat yang artinya tidak ada opini yang diberikan atas laporan keuangan tersebut. Terlepas dari apakah WTP termasuk jenis opini atau tidak, status ini adalah status terendah dalam hal kewajaran laporan keuangan.

Beberapa perbedaan utama antara sektor publik dan sektor swasta yang membuat karakter kedua laporan tersebut berbeda antara lain:

- a. Orientasi pada laba

Sektor swasta sangat menekankan pada perolehan laba, sehingga prinsip layak temu biaya-pendapatan (*matching-cost against revenue principle*) sangat ditekankan dalam operasionalnya. Setiap sen uang yang dibelanjakan harus dapat ditandingkan dengan pendapatan yang diterima akibat belanja tersebut. Selisih antara biaya operasional dengan pendapatan operasional inilah laba yang sangat dicari sektor swasta. Karena tujuan utamanya adalah mencari laba yang identik dengan pendapatan, maka sektor swasta sering disebut dengan entitas pusat pendapatan (*revenue center entity*).

Sektor publik tidak terlalu menekankan prinsip ini, karena seringkali biaya yang dikeluarkan pemerintah untuk menyediakan layanannya tidak selalu langsung menghasilkan pendapatan. Misalnya belanja investasi pembangunan

jalan dan jembatan untuk daerah terpencil, tidak serta merta pemerintah mengharapkan pendapatan dari belanja tersebut. Karena sektor publik/pemerintahan berorientasi pelayanan yang identik dengan belanja, maka sektor publik sering disebut dengan entitas pusat belanja/biaya (*cost center entity*).

b. Transfer pendapatan antar pemerintah

Sektor publik secara substansi terdiri dari tiga lingkup pemerintahan yaitu pemerintah pusat, pemerintah provinsi, dan pemerintah kabupaten/kota. Pemerintah yang lebih luas cakupannya memberi arahan pada pemerintah yang cakupannya lebih sempit. Selain itu, ada sistem transfer pendapatan/subsidi antar pemerintah yang tidak lazim di sektor swasta.

Pemerintah dengan pendapatan pajak atau bukan pajak lebih besar akan menyubsidi/mentransfer pendapatan ke pemerintah yang lebih sedikit pendapatannya. Misalnya pemerintah pusat akan memberikan transfer pendapatan berupa Dana Bagi Hasil Pajak (DBH Pajak) dan Dana Bagi Hasil Sumber Daya Alam (DBH SDA) kepada pemerintah provinsi, kabupaten, kota yang menjadi daerah penghasil SDA tersebut atau daerah lainnya yang berhak sesuai peraturan perundangan. Pemerintah pusat juga memberikan dana alokasi umum (DAU) kepada pemerintah provinsi, kabupaten, kota dengan memperhatikan karakteristik tertentu seperti kesenjangan fiskal/*fiscal gap* pada tiap-tiap daerah penerima DAU tersebut. Pemerintah pusat juga memberikan dana desa yang juga menjadi bukti adanya transfer pendapatan pemerintah pusat ke desa sebagai kesatuan masyarakat hukum. Pemerintah provinsi akan memberikan transfer kepada pemerintah kabupaten atau kota yang berasal dari penerimaan pajak provinsi antara lain: Pajak Kendaraan Bermotor (PKB), Bea Balik Nama Kendaraan Bermotor (BBNKB), Pajak Bahan Bakar Kendaraan Bermotor (PBBKB), Pajak Rokok, dan Pajak Air Permukaan.

Transfer pendapatan antar pemerintah ini juga menjadi pengeluaran pemerintah yang tidak memenuhi prinsip layak temu biaya-pendapatan (*matching-cost against revenue principle*) yang sangat diagungkan dalam ranah swasta. Pemerintah yang melakukan transfer tidak akan menerima pendapatan secara langsung dari setiap pengeluaran transfer yang mereka lakukan.



Pengeluaran transfer ini bertujuan untuk terciptanya pemerataan pendapatan atau distribusi pendapatan (*income redistribution*) dari golongan kaya ke golongan miskin untuk menghindari terjadinya keresahan dan kecemburuan sosial akibat adanya ketimpangan pendapatan yang berpotensi menimbulkan gangguan pada stabilitas nasional [30].

Prinsip distribusi pendapatan tersebut diadopsi oleh Pemerintah dari beberapa negara maju seperti Amerika Serikat, Inggris, dan Belanda yang dikenal dengan “negara kemakmuran” (*welfare state*). Negara-negara tersebut membuat alat-alat kebijakan fiskal untuk tujuan pemerataan pendapatan yang biasanya dilakukan dengan cara:

- 1) Membuat dan melaksanakan sistem pajak yang bersifat progresif. Contoh penerapan kebijakan ini di Indonesia adalah tarif pajak penghasilan (PPh) yang progresif selaju dengan kenaikan pendapatan wajib pajak, Pajak Kendaraan Bermotor (PKB) progresif sesuai dengan jumlah kepemilikan atas pajak bermotor, cukai hasil tembakau dan Pajak Rokok selaju dengan konsumsi atas tembakau atau rokok. Kebijakan ini adalah “pengambilan paksa” pendapatan dari golongan kaya yang dilakukan oleh negara. Orang yang berpenghasilan besar, orang yang memiliki kendaraan bermotor, orang yang mengonsumsi rokok dianggap lebih kaya dari orang yang penghasilannya rendah/orang dengan penghasilan tidak kena pajak (PTKP), orang yang tidak punya kendaraan bermotor, atau orang yang tidak mengonsumsi rokok, sehingga wajar jika negara mengambil sebagian pendapatan mereka untuk dialokasikan bagi warga negara lain yang lebih miskin atau membutuhkan.
- 2) Membuat dan melaksanakan kebijakan belanja negara yang bersifat membantu golongan miskin. Belanja negara seperti ini sering disebut dengan “pembelanjaan kebajikan” (*welfare expenditure*). Berbagai belanja pemerintah seperti subsidi listrik, subsidi gas 3 kg, dan subsidi pupuk adalah bukti nyata belanja negara untuk membantu golongan miskin ini. Dana *earmarking* adalah contoh lain dari belanja kebajikan ini, contoh dari dana *earmarking* ini adalah:

- a) Kewajiban untuk pengalokasian dana 10% dari Pajak Kendaraan Bermotor (PKB) bagi pemeliharaan dan pembangunan jalan, serta peningkatan sarana transportasi umum.
  - b) Kewajiban untuk pengalokasian dana 50% dari Pajak Rokok untuk mendanai pelayanan kesehatan dan penegakan hukum.
  - c) Pengalokasian sebagian penerimaan Pajak Penerangan Jalan (PPJ) digunakan untuk penyediaan penerangan jalan.
- 3) Pemberian bantuan untuk peningkatan kualitas sumber daya manusia (SDM). Indonesia memberikan bantuan untuk hal ini misalnya dalam bentuk Bantuan Operasional Sekolah (BOS), beasiswa pendidikan melalui Lembaga Pengelola Dana Pendidikan (LPDP), beasiswa pendidikan melalui sekolah atau Perguruan Tinggi Kedinasan (PTK) seperti Politeknik Keuangan Negara STAN, Sekolah Tinggi Ilmu Statistik (STIS), Sekolah Tinggi Teknologi Meteorologi Klimatologi dan Geofisika (STTMKG) dan sebagainya.
- c. Pengaruh proses politik
- Laporan keuangan sektor publik sangat dipengaruhi oleh proses politik. Definisi politik sesuai Kamus Besar Bahasa Indonesia (KBBI) adalah: *“segala urusan dan tindakan (kebijakan, siasat, dan sebagainya) mengenai pemerintahan negara atau terhadap negara lain”*. Sedangkan sektor swasta seringkali tidak mau pusing dengan urusan yang berbau kenegaraan. Keputusan untuk mengambil atau melunasi utang, memberikan atau mencabut subsidi di sektor publik semua harus melibatkan proses politik melalui persetujuan legislatif (DPR/D). Jadi penganggaran di sektor swasta lebih fleksibel dibandingkan sektor publik.
- d. Hubungan antara pembayaran pajak dan pelayanan pemerintah
- Sektor publik dhi. Pemerintah berhak memungut pajak sesuai kewenangan yang diberikan peraturan perundangan. Jumlah pajak yang dipungut pemerintah tidak selalu terkait langsung dengan pelayanan yang diberikan pemerintah. Kondisi ini tidak ada di sektor swasta, karena pendapatan yang mereka peroleh dihasilkan dari adanya barang/jasa yang mereka jual kepada pelanggan.

## 2.4 Data Laporan Keuangan

Laporan keuangan sektor publik terdiri dari laporan anggaran (*budgetary reports*), laporan finansial (*financial reports*), dan Catatan atas Laporan Keuangan (CaLK). Dalam basis akuntansi *cash toward accrual* (CTA) jenis laporan keuangan ada 4 yaitu:

a. Laporan Realisasi Anggaran/LRA yang termasuk laporan anggaran

Laporan Realisasi Anggaran menyajikan perbandingan antara data anggaran yang dikelola pemerintah dan data realisasinya untuk tahun tertentu, antara lain:

- 1) Anggaran dan realisasi pendapatan;
- 2) Anggaran dan realisasi belanja;
- 3) Anggaran dan realisasi transfer;
- 4) Anggaran dan realisasi pembiayaan.

b. Neraca yang termasuk laporan finansial

Neraca menggambarkan posisi keuangan suatu entitas yang meliputi data aset (aset lancar, investasi jangka panjang, aset tetap, aset tak berwujud dan sebagainya), data utang/kewajiban baik kewajiban jangka pendek maupun jangka panjang, serta ekuitas yang merupakan kekayaan bersih pemerintah (selisih antara aset dan kewajiban pemerintah).

c. Laporan Arus Kas/LAK yang termasuk laporan finansial

Laporan Arus Kas menyajikan data dan informasi terkait saldo awal, penerimaan, pengeluaran dan saldo akhir kas pemerintah selama kurun waktu tertentu. Penerimaan dan pengeluaran kas dibagi ke dalam aktivitas operasi, investasi aset nonkeuangan, pembiayaan, dan nonanggaran. Sedangkan dalam basis akrual murni aktivitas penerimaan dan pengeluaran kas dibagi dalam kelompok aktivitas operasi, investasi, pendanaan, dan transitoris.

d. CaLK

CaLK merupakan laporan yang berisi rincian dan penjelasan lebih lanjut dari laporan-laporan lainnya. Laporan ini merupakan satu kesatuan yang tidak terpisahkan dari laporan anggaran dan laporan finansial.

Jenis laporan keuangan dalam basis akuntansi akrual murni ada 7 laporan sebagai berikut:

- a. Laporan Anggaran terdiri dari 2 jenis yaitu: Laporan Realisasi Anggaran (LRA) dan Laporan Perubahan Saldo Anggaran Lebih (LP SAL).
- b. Laporan Finansial terdiri dari 4 jenis yaitu: Neraca, Laporan Operasional (LO), Laporan Perubahan Ekuitas (LPE), dan Laporan Arus Kas (LAK).
- c. Catatan atas Laporan Keuangan (CaLK).

Namun karena penggunaan basis akrual murni belum diterapkan oleh semua pemerintah daerah bahkan hingga laporan keuangan periode tahun anggaran 2014, sehingga penelitian ini menggunakan data keuangan yang berasal dari LRA dan Neraca, karena kedua laporan tersebut wajib ada baik dalam basis akrual murni maupun CTA. Data-data keuangan tersebut antara lain realisasi belanja, realisasi pendapatan, realisasi transfer, realisasi pembiayaan, posisi aset, posisi utang, dan posisi ekuitas.

## **2.5 Penambangan Data (*Data Mining/Knowledge Discovery in Database*)**

Penelitian [31] menyatakan bahwa *knowledge discovery in database* (KDD) mengacu pada keseluruhan proses untuk mengungkap pengetahuan yang berguna dari data, sedangkan penambangan data/*data Mining (DM)* merupakan salah satu langkah dalam proses KDD itu sendiri. Penambangan data merupakan pengaplikasian suatu algoritme khusus untuk mengekstrak pola dari suatu data. Adapun rincian langkah-langkah yang ada dalam KDD sebagai berikut:

- a. Pemilihan data, pada tahap ini dipilih data yang dibutuhkan untuk proses penambangan data.
- b. Pra-pemrosesan, pada tahap ini biasanya dilakukan pembersihan data untuk menghilangkan data yang kurang bagus seperti data tidak lengkap dan data yang tidak valid.
- c. Transformasi data, pada tahap ini data diubah ke format yang sesuai dengan kebutuhan teknik penambangan data yang akan digunakan.
- d. Penambangan data, pengaplikasian suatu algoritme khusus untuk mengekstrak pola dari data.

- e. Interpretasi/penarikan kesimpulan, proses untuk menghasilkan pengetahuan dari pola data yang telah didapatkan.

Penambahan data terkait dengan empat kelompok pekerjaan sebagai berikut:

- a. Model prediksi

Model ini memetakan setiap himpunan variabel masukan ke setiap targetnya (variabel keluaran), kemudian menggunakan pola yang didapat untuk memberikan nilai keluaran dari data masukan baru. Jenis model prediksi ada dua macam, yaitu klasifikasi dan regresi.

- b. Analisis kelompok

Model yang digunakan untuk mengelompokkan data-data ke dalam sejumlah kelompok (klaster) berdasarkan kesamaan karakteristik data pada masing-masing kelompok. Data pada satu klaster memiliki tingkat kesamaan karakter yang tinggi dengan data dalam satu klaster dan memiliki tingkat kesamaan karakter yang rendah dengan data pada klaster lain.

- c. Analisis asosiasi

Model yang digunakan untuk menemukan pola hubungan kekuatan antar fitur dalam suatu data. Contoh penggunaan model ini dalam bisnis adalah mencari hubungan antara beberapa barang yang sering dibeli konsumen secara bersamaan (misal: membeli sikat gigi dan pasta gigi) yang dapat digunakan untuk penataan letak barang dagangan di toko.

- d. Deteksi anomali

Model yang digunakan untuk mencari data yang memiliki karakteristik “aneh” dibandingkan dengan karakteristik data lainnya [32].

## **2.6 Metode Tetangga Terdekat/*K-Nearest Neighbors***

Tetangga terdekat (NN) adalah salah satu dari sepuluh metode yang paling banyak digunakan di penambahan data. Metode NN murni diklasifikasikan sebagai *lazy learner*, karena NN menunda proses pelatihan atau bahkan tidak melakukan sama sekali, sampai ada data uji yang ingin diketahui label kelasnya.

ketika ada data uji yang ingin diketahui label kelasnya, maka metode NN baru akan menjalankan menjalankan algoritmanya [25].

*K-Nearest Neighbors* (KNN) adalah metode berdasarkan konsep NN yang paling tua, paling sederhana dan populer namun memiliki kinerja yang baik yang mampu menyamai metode lain yang lebih rumit seperti PNN [24], [26], [27]. Algoritma klasifikasi KNN dijalankan berdasar konsep kedekatan atau jarak data dengan data lainnya. Nilai-K dalam KNN menyatakan jumlah data terdekat yang digunakan untuk penentuan kelas dalam klasifikasi.

Ada beberapa isu penting yang mempengaruhi kinerja KNN [25], termasuk pemilihan nilai-K. Jika nilai-K terlalu kecil, maka hasil prediksi akan sensitif terhadap gangguan/*noise*. Di sisi lain, jika nilai-K terlalu besar, maka ada risiko bahwa data yang dipilih tidak relevan karena mayoritas data yang dipilih bisa dari data yang berlainan kelas atau terlalu jauh dari data yang ingin dikenali label kelasnya. Isu lain yang berkaitan dengan K-NN adalah tentang voting suara terbanyak dalam klasifikasi. Jika nilai-K ganjil, risiko terjadi dua kelas atau lebih yang memperoleh jumlah suara sama sangat kecil. Namun ketika nilai-K genap, ada risiko dua kelas atau lebih memiliki jumlah suara yang sama, sehingga algoritma akan memilih kelas secara acak. Selain itu, KNN juga tidak memberikan bobot pengaruh pada data yang dipilih seperti pada jaringan syaraf tiruan/*Neural Network*, untuk mengatasi kendala-kendala tersebut, algoritma KNN biasanya diaplikasikan dengan validasi 10-lipatan (*10-folds validation*), menggunakan aturan jarak terdekat (*tie-break rules*) untuk mengatasi stagnasi karena jumlah suara yang sama kuat, dan pemberian bobot pengaruh berdasarkan jarak untuk memberikan pengaruh yang lebih kecil terhadap data yang terletak jauh dari data uji.

### **2.6.1 Konsep Kedekatan**

Secara umum metode klasifikasi atau metode *clustering* menggunakan kuantitas yang disebut kedekatan/*proximity*. Kedekatan secara konsep dibagi dua yaitu kemiripan/*similarity* dan ketidakmiripan/*dissimilarity*. Secara teoritis kedekatan diukur dengan mencari jarak antara data pertama terhadap data kedua. Semakin dekat jarak antar data, maka dapat dikatakan semakin besar kemiripan atau semakin kecil ketidakmiripan antar data tersebut. Sebaliknya semakin jauh jarak

antar data maka dapat dikatakan bahwa semakin kecil kemiripan atau semakin besar ketidakmiripan antar data tersebut. Ukuran ketidakmiripan yang sering dipakai adalah jarak Euclidean, Manhattan/City Block, Minkowsky, dan Chebisev. Penelitian ini hanya menggunakan konsep kedekatan menggunakan Euclidean dan Manhattan karena dua konsep jarak tersebut yang paling sering dipakai. Adapun formulasi jarak Euclidean disajikan pada Persamaan (2.1) dan Manhattan pada Persamaan (2.2) sebagai berikut:

$$D(\bar{x}, \bar{y}) = \sqrt{\sum_{j=1}^N (x_j - y_j)^2} \quad (2.1)$$

$$D(\bar{x}, \bar{y}) = \sum_{j=1}^N |x_j - y_j| \quad (2.2)$$

Dimana:

$D(\bar{x}, \bar{y})$  = jarak antara data x dan data y

$x_j$  = data latih

$y_j$  = data uji

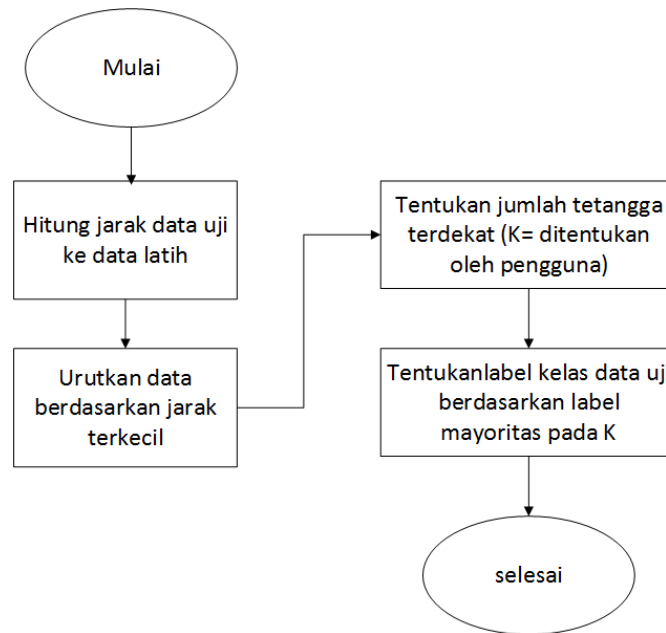
N = jumlah fitur atau variabel

### 2.6.2 Tahapan Algoritma KNN

Tahapan algoritma KNN dapat diuraikan sebagai berikut:

- a. Cari jarak (bisa menggunakan alternatif jarak yang lazim dipakai) antara data latih dan data uji yang ingin diketahui label kelasnya;
- b. Urutkan data latih berdasarkan jarak terkecil dari data uji;
- c. Tentukan jumlah K (tetangga terdekat) yang akan digunakan untuk penentuan label kelas. Penentuan jumlah/nilai-K ini merupakan kunci kesuksesan klasifikasi dengan metode KNN. Nilai-K ini ditentukan oleh pengguna baik dengan metode coba-coba atau menggunakan metode lainnya seperti metode optimasi.
- d. Tentukan label kelas data uji berdasarkan label dengan suara mayoritas sesuai langkah (c).

Tahapan algoritma KNN tersebut diilustrasikan pada Gambar 2.1 berikut:



Gambar 2.1 Diagram alir tahapan algoritma KNN (adaptasi dari [33])

## 2.7 Metode Evaluasi Klasifikator

Evaluasi klasifikator dilakukan untuk mengetahui klasifikator mana yang memberikan hasil terbaik. Agar evaluasi bebas dari masalah hasil yang terlalu bagus/*overfitting*, maka biasanya yang dipakai dalam pengambilan keputusan adalah hasil klasifikasi terhadap data uji bukan klasifikasi terhadap data latih. Penelitian ini menggunakan dua dari sekian banyak metode evaluasi klasifikator yaitu Metode *Hold-Out* dan Metode Validasi Silang 10-Lipatan, adapun penjelasan dari kedua metode tersebut sebagai berikut:

### 2.7.1 *Hold-Out*

Metode *Hold-Out* (HO) [25] memecah set data yang sudah diketahui label kelasnya menjadi dua bagian terpisah yaitu set data latih dan set data uji. Model klasifikasi dibangun menggunakan set data latih sedangkan pengukuran kinerja modelnya menggunakan set data uji. Besaran pembagian set data ini bersifat diskrit (bilangan bulat, misalnya HO 10% berarti pembagian set data 10% untuk data uji dan 90% untuk data latih).



Metode ini merupakan metode yang paling sederhana dengan beberapa keterbatasan:

- a. Jumlah data latih menjadi berkurang, karena sebagian digunakan untuk data uji, sehingga model yang dibangun mungkin tidak sebagus ketika semua data digunakan sebagai data latih;
- b. Model yang dibangun sangat tergantung pada komposisi pemecahan set data latih dan set data uji. Set data latih yang semakin sedikit berisiko menimbulkan variansi dari model, sedangkan set data latih yang semakin banyak menimbulkan risiko tidak dapat dipercayanya akurasi klasifikasi atas data uji (*non-reliable*);
- c. Set data latih dan set data uji tidak lagi bebas satu sama lain, kelebihan jumlah perwakilan pada data uji menimbulkan kekurangan jumlah perwakilan di data latih dan sebaliknya.

### 2.7.2 Validasi Silang 10-Lipatan

Bentuk umum dari metode ini disebut Validasi Silang k-Lipatan (*k-Folds Cross Validation*). Metode ini memecah set data menjadi k subset dengan ukuran yang sama [25]. Setiap kali proses dilakukan, satu subset berperan sebagai data uji sedangkan subset yang lain menjadi data latih. Prosedur tersebut diulang sebanyak k-kali sehingga setiap data berkesempatan menjadi data uji tepat satu kali dan menjadi data latih sebanyak k-1 kali. Total akurasi didapatkan dengan menjumlahkan semua akurasi pada setiap proses k.

Validasi Silang 10-Lipatan membagi set data menjadi 10 subset data. Prosedur pengulangan 10 kali membuat setiap data berkesempatan menjadi data uji sebanyak satu kali dan data latih sebanyak sembilan kali. Metode ini lebih baik dalam membangun model dibandingkan metode HO [25].

## 2.8 Uji Wilcoxon Mann Whitney

Uji Wilcoxon Mann Whitney (WMW) [34] merupakan uji nonparametrik yang digunakan untuk menentukan apakah dua sampel ( $n=2$ ) yang saling bebas (*independen*) dapat dianggap memiliki perbedaan yang signifikan atau tidak. Kriteria dua atau lebih sampel dikatakan berpasangan atau terikat

(*paired/dependen*) adalah jika data tersebut berasal dari individu yang sama, baik karena proses pencocokan (*matching*), pengukuran berulang atau karena desain penelitian silang (*crossover*). Dua atau lebih sampel dikatakan bebas (*independen*) jika berasal dari subyek/individu yang berbeda tanpa prosedur pencocokan/*matching* [35]. Metode ini menggunakan peringkat data untuk perhitungannya, bukan data asli observasi sehingga proses perhitungan menjadi lebih sederhana. Metode ini diusulkan untuk proses seleksi fitur pada klasifikasi dengan dua label kelas dengan asumsi bahwa fitur yang memiliki perbedaan signifikan antar kelas akan mampu memprediksi kelas dari suatu data lebih baik daripada fitur yang tidak memiliki perbedaan signifikan antar kelas.

Persamaan untuk Uji Wilcoxon Mann Whitney [36] jika tidak terdapat nilai observasi yang sama disajikan pada Persamaan (2.3) berikut:

$$Z = \frac{(W_x \pm 0,5) - \frac{1}{2}m(N+1)}{\sqrt{\frac{mn(N+1)}{12}}} \quad (2.3)$$

Dimana:

Z = nilai dari Uji Wilcoxon Mann Whitney

m = ukuran sampel kelompok yang kecil

n = ukuran sampel kelompok yang besar

N = m+n, total data observasi

$W_x$  = jumlah peringkat (*ranking*) pada kelompok sampel yang kecil

Jika terdapat data pengamatan dengan nilai yang sama [36], maka koreksi untuk Persamaan (2.3) disajikan pada Persamaan (2.4) berikut:

$$Z_c = \frac{(W_x \pm 0,5) - \frac{1}{2}m(N+1)}{\sqrt{\frac{mn}{N(N-1)} \left( \frac{N^3 - N}{12} - T \right)}} \quad (2.4)$$

Dimana:

$Z_c$  = nilai koreksi dari Uji Wilcoxon Mann Whitney

$T$  =  $\frac{1}{12} (t^3 - t)$

$t$  = banyaknya pengamatan yang sama pada skor tertentu

Ilustrasi Uji Wilcoxon Mann Whitney untuk menguji apakah ada perbedaan yang signifikan antara umur peserta Kelas A dan B disajikan pada Tabel 2.2 berikut:

Tabel 2.2 Data Umur Peserta pada Tiap Kelas

No	Kelas A	Kelas B
1	46	44
2	32	31
3	42	25
4	45	22
5	37	30
6	44	30
7	38	32
8	47	19
9	49	40
10	41	
<b>Rerata</b>	<b>42,10</b>	<b>30,33</b>
<b>Variansi</b>	<b>27,21</b>	<b>63,75</b>

Tahap selanjutnya adalah mencari peringkat data secara keseluruhan tanpa melihat dari kelas mana data tersebut berasal. Apabila terdapat data dengan nilai yang sama maka peringkat untuk data-data tersebut menggunakan nilai rerata

peringkat atau  $\frac{\sum \text{peringkat}}{\sum \text{data yang sama}}$ , misalkan data nomor urut/peringkat 1 dan 2

sama, maka peringkat bagi kedua data tersebut adalah  $\frac{(1+2)}{2} = 1,5$ . Kriteria untuk

menerima atau menolak  $H_0$  adalah jika nilai  $p\text{-value} \leq \alpha$  maka  $H_0$  ditolak, sedangkan jika  $p\text{-value} > \alpha$  maka  $H_0$  diterima. Hipotesis untuk data ilustrasi tersebut adalah:

$H_0$  = tidak ada perbedaan yang signifikan antara umur peserta di Kelas A dan B

$H_1$  = ada perbedaan yang signifikan antara umur peserta di Kelas A dan B

Data ilustrasi mengandung nilai observasi yang sama sehingga Persamaan (2.4) yang akan digunakan dalam penentuan penolakan atau penerimaan  $H_0$ . Perhitungan nilai T untuk data dengan skor sama disajikan pada Tabel 2.3 berikut:

Tabel 2.3 Skor T untuk Data yang Sama pada Uji Wilcoxon Mann Whitney

Skor	Frekuensi (t)	T
32	2	0,5
44	2	0,5
30	2	0,5
	$\Sigma T$	1,5

Berdasarkan hasil perhitungan Uji Wilcoxon Mann Whitney diketahui bahwa variabel/fitur umur memiliki perbedaan yang signifikan antara Kelas A dan B, sehingga jika dilakukan proses seleksi fitur untuk klasifikasi, maka fitur umur adalah fitur yang dianggap signifikan dalam klasifikasi. Hasil pemeringkatan data umur peserta beserta perhitungan Uji Wilcoxon Mann Whitney serta keputusan untuk menolak atau menerima  $H_0$  pada tingkat signifikansi 1% ( $\alpha = 0,01$ ) secara detail disajikan pada Tabel 2.4 berikut:

Tabel 2.4 Hasil Pemeringkatan Data Umur dan Keputusan atas  $H_0$

No	Kelas A	Kelas B
1	17	14,5
2	7,5	6
3	13	3
4	16	2
5	9	4,5
6	14,5	4,5
7	10	7,5
8	18	1
9	19	11
10	12	
Hasil penjumlahan peringkat ( $\Sigma R$ )	136	54
m	9	
n	10	
N	19	
$W_x$	54	
$Z_c$	-2,90	
<i>p-value</i>	0,0019	
$\alpha$	0,01	

No	Kelas A	Kelas B
Keputusan	Terima H1	

## 2.9 Uji Kruskal-Wallis

Uji Kruskal-Wallis (KW) menurut [37] merupakan uji nonparametrik yang digunakan untuk menentukan apakah lebih dari dua sampel ( $n > 2$ ) yang saling bebas (*independen*) dapat dianggap berasal dari populasi yang sama atau tidak. Metode ini juga menggunakan peringkat data untuk perhitungannya, bukan data asli observasi sehingga proses perhitungan menjadi lebih sederhana. Metode ini diusulkan untuk proses seleksi fitur pada klasifikasi dengan empat label kelas ( $n > 2$ ) dengan asumsi bahwa fitur yang berasal dari populasi yang berbeda tiap kelasnya akan mampu memprediksi kelas dari suatu data lebih baik daripada fitur yang berasal dari kelas dengan populasi yang sama.

Persamaan untuk Uji Kruskal-Wallis jika tidak terdapat nilai yang sama sebagaimana dinyatakan oleh [37] disajikan pada Persamaan (2.5) berikut:

$$H = \frac{12}{N(N+1)} \sum_{i=1}^C \frac{R_i^2}{n_i} - 3(N+1) \quad (2.5)$$

Dimana:

H = nilai dari Uji Kruskal-Wallis

C = jumlah kelas

$n_i$  = jumlah data observasi pada kelas ke-i

N =  $\sum n_i$ , jumlah data observasi pada keseluruhan kelas

$R_i$  = jumlah peringkat (*ranking*) pada kelas ke-i

Jika terdapat data pengamatan dengan nilai yang sama maka koreksi untuk Persamaan (2.5) sebagaimana dinyatakan [37] adalah:

$$H_c = \frac{H}{1 - \frac{\sum T}{N^3 - N}} \quad (2.6)$$

Dimana:

$H_c$  = nilai koreksi dari Uji Kruskal-Wallis jika terdapat data sama

T =  $t^3 - t$

t = banyaknya pengamatan yang sama pada skor tertentu

Ilustrasi Uji Kruskal-Wallis pada tiga kelas A, B, dan C berdasarkan fitur umur peserta (dalam satuan tahun) tiap kelas disajikan pada Tabel 2.5 berikut:

Tabel 2.5 Data Umur Peserta pada Tiap Kelas

No	Kelas A	Kelas B	Kelas C
1	46	44	26
2	32	31	49
3	42	25	33
4	45	22	19
5	37	30	31
6	44	30	38
7	38	32	44
8	47	19	50
9	49	40	
10	41		
<b>Rerata</b>	<b>42,10</b>	<b>30,33</b>	<b>36,25</b>
<b>Variansi</b>	<b>27,21</b>	<b>63,75</b>	<b>122,21</b>

Tahap selanjutnya adalah mencari peringkat data secara keseluruhan tanpa melihat dari kelas mana data tersebut berasal. Apabila terdapat data dengan nilai yang sama maka peringkat bagi data-data tersebut menggunakan nilai rerata

peringkat atau  $\frac{\sum \text{peringkat}}{\sum \text{data yang sama}}$ , misalkan data nomor urut/peringkat 1 dan 2

sama, maka peringkat bagi kedua data tersebut adalah  $\frac{(1+2)}{2} = 1,5$ . Kriteria untuk

menerima atau menolak  $H_0$  adalah jika nilai  $H$  hitung  $\geq H$  tabel maka  $H_0$  ditolak, sedangkan jika  $H$  hitung  $< H$  tabel maka  $H_0$  diterima. Hipotesis untuk data ilustrasi tersebut adalah:

$H_0$  = data sampel berasal dari populasi yang sama

$H_1$  = data sampel berasal dari populasi yang berbeda

Data ilustrasi mengandung nilai observasi yang sama sehingga Persamaan (2.6) yang akan digunakan dalam penentuan penolakan atau penerimaan  $H_0$ . Perhitungan nilai  $T$  untuk data dengan skor sama disajikan pada Tabel 2.6 berikut:

Tabel 2.6 Skor T untuk Data yang Sama pada Uji Kruskal-Wallis

Skor	Frekuensi (t)	T
44	3	24
32	2	6
38	2	6
49	2	6
31	2	6
30	2	6
19	2	6
	$\Sigma T$	<b>60</b>

Berdasarkan hasil perhitungan Uji Kruskal-Wallis diketahui bahwa variabel/fitur umur memiliki data sampel yang berasal dari populasi yang sama sehingga jika dilakukan proses seleksi fitur untuk klasifikasi, maka fitur umur adalah fitur yang tidak signifikan digunakan dalam klasifikasi. Hasil pemeringkatan data umur peserta beserta perhitungan Uji Kruskal-Wallis serta keputusan untuk menolak atau menerima  $H_0$  pada tingkat signifikansi 1% ( $\alpha = 0,01$ ) secara detail disajikan pada Tabel 2.7 berikut:

Tabel 2.7 Hasil Pemeringkatan Data Umur dan Keputusan atas  $H_0$ 

No	Kelas A	Kelas B	Kelas C	Total
1	23,0	20,0	5,0	
2	10,5	8,5	25,5	
3	18,0	4,0	12,0	
4	22,0	3,0	1,5	
5	13,0	6,5	8,5	
6	20,0	6,5	14,5	
7	14,5	10,5	20,0	
8	24,0	1,5	27,0	
9	25,5	16,0		
10	17,0			
Hasil penjumlahan peringkat ( $R_i$ )	187,5	76,5	114,0	
Jumlah data observasi ( $n_i$ )	10	9	8	27
$\frac{R^2}{n_i}$	3.515,63	650,25	1.624,50	5.790,38

Tabel 2.7 Hasil Pemeringkatan Data Umur dan Keputusan atas H0 (Lanjutan)

No	Kelas A	Kelas B	Kelas C	Total
H				7,9107
$\sum T$				60
N3-N				19.656
Hc				7,9349
Derajat kebebasan (df)				2
Nilai H tabel				9,2103
<b>Keputusan</b>				<b>Terima H0</b>

Uji lanjutan diperlukan untuk memperkuat pengambilan keputusan atas signifikan atau tidaknya suatu fitur untuk digunakan dalam proses klasifikasi. Uji ini dilakukan jika hasil Uji Kruskal-Wallis menunjukkan ada perbedaan populasi diantara kelompok yang diuji. Uji lanjutan ini dapat dilakukan dengan Uji Wilcoxon Mann Whitney atau Uji Beda Nyata Jujur Tukey (Tukey *Honest Significant Difference*/Tukey *HSD*). Kriteria Uji Tukey *HSD* [38] untuk menyatakan Grup A dan B berbeda secara signifikan adalah jika:

$$|\bar{A} - \bar{B}| > \frac{Q_{\alpha, p, dfe}}{\sqrt{2}} \times \sqrt{KTG \left( \frac{1}{n_i} + \frac{1}{n_j} \right)} \quad (2.7)$$

Dimana:

$|\bar{A} - \bar{B}|$  = nilai absolut dari rerata Grup A – rerata Grup B

$Q_{\alpha, p, dfe}$  = nilai batas Uji Tukey dengan kriteria  $\alpha$ , p, dan derajat kebebasan ( $dfe$ )

$\alpha$  = tingkat akurasi nyata penelitian

p = banyaknya grup atau perlakuan

$dfe$  = derajat kebebasan model

KTG = kuadrat tengah akurasi (MSE)

$n_{i,j}$  = banyaknya data pengamatan pada grup yang ditandingkan

Uji lanjutan tersebut akan memberikan gambaran yang lebih jelas mengenai kelas mana saja yang memiliki perbedaan signifikan/berasal dari populasi yang berbeda. Penelitian ini menggunakan Uji Kruskal-Wallis yang dilanjutkan dengan uji



lanjutan Tukey *HSD* dan Wilcoxon Mann Whitney untuk penentuan fitur signifikan. Perbandingan kinerja antara seleksi fitur menggunakan uji lanjutan Tukey *HSD* dan Wilcoxon Mann Whitney dilakukan untuk mengetahui metode seleksi fitur mana yang lebih baik.

## 2.10 Uji Friedman

Uji Friedman [39] merupakan uji nonparametrik lebih dari dua sampel ( $n > 2$ ) yang tidak saling bebas (*dependen*) atau berpasangan (*paired*) memiliki pengaruh signifikan yang sama atau tidak. Uji ini dilakukan dengan menggunakan tabel dua arah. Baris sebagai blok/kategori yang mendapatkan perlakuan dan kolom sebagai variabel yang diisolasi atau perlakuan yang diuji pengaruhnya terhadap blok. Uji ini menggunakan peringkat data dalam perhitungannya. Uji ini bertujuan menguji apakah perbedaan antara rerata peringkat tiap kolom dapat dianggap berbeda secara signifikan atau tidak.

Persamaan untuk Uji Friedman jika tidak terdapat nilai yang sama sebagaimana dinyatakan oleh [39] disajikan pada Persamaan (2.8) berikut:

$$\chi_r^2 \text{ atau } S = \frac{12}{np(p+1)} \sum_{j=1}^p \left( \sum_{i=1}^n r_{ij} \right)^2 - 3n(p+1) \quad (2.8)$$

Dimana:

$\chi_r^2$  atau  $S$  = nilai dari Uji Friedman  
 $n$  = jumlah blok percobaan  
 $p$  = jumlah perlakuan percobaan  
 $r_{ij}$  = jumlah peringkat (*ranking*) pada kelas ke- $j$

Jika terdapat data pengamatan dengan nilai yang sama [40], maka koreksi untuk Persamaan (2.8) adalah:

$$S_c = \frac{S}{1 - \frac{\sum T}{n(p^3 - p)}} \quad (2.9)$$

Dimana:

$S_c$  = nilai koreksi dari Uji Friedman jika terdapat data sama

$T = t^3 - t$   
 $t =$  banyaknya pengamatan yang sama pada skor tertentu

Uji lanjutan (*post hoc test*) dari Uji Friedman dengan hasil “menolak  $H_0$ ” untuk mengetahui perbandingan grup/perlakuan mana yang signifikan pada dua sampel yang tidak saling bebas atau berpasangan dapat dilakukan dengan Uji Beda Nyata Jujur Tukey (Tukey *Honest Significant Difference*/Tukey *HSD*) atau Uji Wilcoxon Peringkat Bertanda.

### 2.11 Uji Wilcoxon Peringkat Bertanda (*Wilcoxon Sign Rank*)

Uji Peringkat Bertanda Wilcoxon (WSR) [34] merupakan uji nonparametrik yang digunakan untuk menentukan apakah dua sampel ( $n=2$ ) yang tidak saling bebas (*dependen*) atau berpasangan (*paired*) memiliki perbedaan yang signifikan atau tidak. Uji ini dilakukan sebagai uji lanjutan dari Uji Friedman untuk menilai apakah ada perbedaan signifikan antar tiap skema (fitur, normalisasi, jarak, dan nilai-K). Persamaan untuk Uji Wilcoxon Peringkat Bertanda [41] disajikan pada Persamaan (2.10) berikut:

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \quad (2.10)$$

Dimana:

$Z$  = nilai dari Uji Wilcoxon Peringkat Bertanda  
 $T$  = jumlah peringkat dengan tanda lebih sedikit  
 $n$  = ukuran data yang tidak memiliki selisih nol

### 2.12 *F-Measure*

*F-Measure* atau *F<sub>β</sub>-Measure* [42] merupakan metrik evaluasi hasil kombinasi dari *recall*/sensitivitas dan *precision/positive predictive value* yang cukup populer karena efektivitasnya dalam pencarian kembali informasi (*information retrieval*), bahkan jika terdapat masalah ketidakseimbangan data. Metrik ini dipengaruhi parameter  $\beta$  yang memiliki jangkauan nilai mulai dari 0

sampai tak hingga (*infinity*) yang berguna untuk mengontrol pengaruh dari *recall* dan *precision* secara terpisah. Persamaan (2.11) menyajikan perhitungan *F-Measure* menurut [43]:

$$F_{\beta} = \frac{(\beta^2 + 1) \times P \times R}{\beta^2 \times P + R} \quad (2.11)$$

Dimana:

- $F_{\beta}$  = F-Measure pada parameter  $\beta$
- $\beta$  = Relativitas pengaruh *Recall* atas Presisi
- $P$  = Presisi
- $R$  = *Recall/True Positive Rate/Sensitivity*

Ketika *recall* dan presisi menurut [43] memiliki bobot pengaruh yang sama kuat, maka nilai  $\beta = 1$ . Jika *recall* memiliki pengaruh separuh dari presisi, maka nilai  $\beta = 0,5$ . Jika *recall* memiliki pengaruh dua kali dari presisi, maka nilai  $\beta = 2$ . Nilai  $\beta$  yang paling sering digunakan adalah 1, karena metrik *recall* dan presisi dianggap memiliki pengaruh yang sama kuat. Sehingga Persamaan (2.11) akan menjadi:

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (2.12)$$

### 2.13 Area di Bawah Kurva ROC/AUC

*Area Under the ROC (Receiver Operating Characteristic) Curve* atau AUC menurut [42] adalah ukuran tunggal suatu kinerja klasifikasi yang berguna untuk penilaian model mana yang lebih baik secara rerata. AUC merupakan suatu angka yang berisi rangkuman mengenai informasi kinerja suatu klasifikator yang mempermudah penilaian model terbaik jika tidak ada kurva ROC yang dominan. Jika F-Measure tidak memperhitungkan metrik kesalahan prediksi (*False Positive Rate/FPR*), maka AUC adalah metrik pengukuran yang menghitung luas bidang trapesium yang terbentuk oleh *Recall/True Positive Rate (TPR)* dan *FPR*. Perhitungan AUC disajikan pada Persamaan (2.13) berikut [42]:

$$AUC = \frac{1 + TPR - FPR}{2} \quad (2.13)$$

Hasil perhitungan AUC digunakan untuk pengambilan simpulan akhir apakah model klasifikasi data laporan keuangan untuk deteksi opini laporan keuangan sektor publik menggunakan fitur modifikasian dari sektor swasta layak diaplikasikan atau tidak. Kriteria untuk penentuan kualitas klasifikasi menurut [44] adalah sebagai berikut:

Tabel 2.8 Interpretasi AUC

No	AUC	Interpretasi
1	0,90 - 1,00	<i>Excellent Classification</i>
2	0,80 - 0,90	<i>Good Classification</i>
3	0,70 - 0,80	<i>Fair Classification</i>
4	0,60 - 0,70	<i>Poor Classification</i>
5	0,50 - 0,60	<i>Failure Classification</i>

Sumber: [44]

Penentuan model dengan kinerja terbaik akan menggunakan rerata harmonik (RH) antara nilai *F-measure* dan AUC dengan persamaan sebagai berikut:

$$RH = \frac{2 \times F \times AUC}{(F + AUC)} \quad (2.14)$$

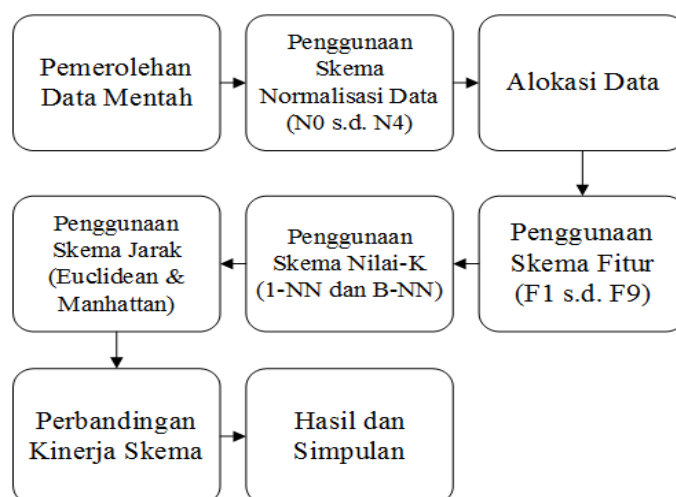
## BAB 3

### METODOLOGI PENELITIAN

Bab ini akan menjelaskan metode dan cara kerja yang akan dipakai dalam penelitian, sehingga akan menghasilkan jawaban atas pertanyaan apakah metode klasifikasi KNN, baik klasifikasi dengan dua kelas maupun klasifikasi dengan empat kelas, cukup bagus untuk menjadi metode alternatif deteksi opini bagi BPK RI. Apabila kinerja klasifikasi menggunakan model ini cukup bagus, maka diharapkan dapat mengurangi risiko kesalahan pemberian opini, mendeteksi terjadinya jual beli opini oleh pemeriksa, dan membantu BPK RI untuk menyusun metode dan langkah pemeriksaan yang mampu memenuhi harapan penugasan pemeriksaan itu sendiri.

#### 3.1 Alur Penelitian

Alur penelitian yang digunakan dalam penelitian ini terdiri dari delapan fase utama: pemerolehan data mentah, penggunaan skema normalisasi data, alokasi data, penggunaan skema fitur, penggunaan skema nilai-K, penggunaan skema jarak, perbandingan kinerja skema, dan terakhir adalah hasil dan simpulan. Alur penelitian tersebut diilustrasikan pada Gambar 3.1 berikut:



Gambar 3.1 Diagram alur penelitian

### 3.2 Pemerolehan Data Mentah

Pada tahap ini dikumpulkan 150 data Laporan Hasil Pemeriksaan atas Laporan Keuangan (LHP LK) Pemerintah Daerah dalam bentuk *softcopy* (pdf) dengan rincian sebagaimana tersaji pada Tabel 3.1 berikut:

Tabel 3.1 Rincian Data LHP LK

No	Uraian	Jumlah LHP LK
1	<b>Zona Wilayah</b> - Barat - Timur	- 82 wilayah barat - 68 wilayah timur
2	Level Pemerintah Daerah	- 28 LK Provinsi - 81 LK Kabupaten - 41 LK Kota
3	Tahun Anggaran (TA) yang diperiksa	- 3 LK TA 2007 - 5 LK TA 2008 - 13 LK TA 2009 - 22 LK TA 2010 - 24 LK TA 2011 - 36 LK TA 2012 - 28 LK TA 2013 - 19 LK TA 2014
4	Opini LK	- 75 WTP - 25 WDP - 25 TW - 25 TMP

Berdasarkan LHP LK tersebut dikumpulkan data kuantitatif sesuai variabel pengukuran laporan keuangan sektor swasta menurut [22] yang telah dimodifikasi untuk sektor publik. Adapun rincian modifikasi variabel tersebut dapat dilihat pada Tabel 3.2.

Tabel 3.2 Modifikasi Variabel Keuangan Sektor Swasta

No	Variabel Laporan Keuangan Sektor Swasta [22]	No	Modifikasi Variabel
1	<i>Debt</i>	1	Hutang
2	<i>Total assets</i>	2	Total Aset
3	<i>Gross profit</i>		*tidak dipakai (pemerintah tidak mencari laba)
4	<i>Net profit</i>	3	SILPA (Sisa Lebih Pembiayaan Anggaran)**
5	<i>Primary business income</i>	4	PAD (pendapatan asli daerah)**
6	<i>Cash and deposits</i>	5	Kas Tunai dan Kas Bank
7	<i>Accounts receivable</i>	6	Piutang
8	<i>Inventory/Primary business income</i>	7	Kas Tunai dan Kas Bank/PAD**
9	<i>Inventory/Total assets</i>	8	Kas Tunai dan Kas Bank/Total Aset
10	<i>Gross profit/Total assets</i>		*tidak dipakai (pemerintah tidak mencari laba)
11	<i>Net profit/Total assets</i>	9	SILPA/Total Aset**
12	<i>Current assets/Total assets</i>	10	Aset Lancar/Total Aset
13	<i>Net profit/Primary business income</i>	11	SILPA/PAD**
14	<i>Accounts receivable/Primary business income</i>	12	Piutang/PAD**
15	<i>Primary business income/Total assets</i>	13	PAD/Total Aset**
16	<i>Current assets/Current liabilities</i>	14	Aset Lancar/Hutang Lancar
17	<i>Primary business income/Fixed assets</i>	15	PAD/Aset Tetap**
18	<i>Cash/Total assets</i>	16	Kas tunai/Total Aset
19	<i>Inventory/Current liabilities</i>	17	Kas Tunai dan Kas Bank/Hutang Lancar
20	<i>Total debt/Total equity</i>	18	Total Hutang/Total Ekuitas

Tabel 3.2 Modifikasi Variabel Keuangan Sektor Swasta (Lanjutan)

No	Variabel Laporan Keuangan Sektor Swasta [22]	No	Modifikasi Variabel
21	<i>Long term debt/Total assets</i>	19	Hutang Jangka Panjang/Total Aset
22	<i>Net profit/Gross profit</i>		*tidak dipakai (pemerintah tidak mencari laba)
23	<i>Total debt/Total assets</i>	20	Total Hutang/Total Aset
24	<i>Total assets/Capital and reserves</i>	21	Total Aset/Total Ekuitas
25	<i>Long term debt/Total capital and reserves</i>	22	Hutang Jangka Panjang/Total Ekuitas
26	<i>Fixed assets/Total assets</i>	23	Aset Tetap/Total Aset
27	<i>Deposits and cash/Current assets</i>	24	Kas Tunai & Kas Bank/Aset Lancar
28	<i>Capitals and reserves/Total debt</i>		*tidak dipakai (sulit dicari padanannya)
29	<i>Accounts receivable/Total assets</i>	25	Piutang/Total Aset
30	<i>Gross profit/Primary business profit</i>		*tidak dipakai (pemerintah tidak mencari laba)
31	<i>Undistributed profit/Net profit</i>		*tidak dipakai (pemerintah tidak mencari laba)
32	<i>Primary business profit/Primary business profit of last year</i>	26	Total Belanja/Total Belanja Tahun Lalu (entitas Belanja bukan entitas pendapatan/ <i>non profit oriented</i> )**
33	<i>Primary business income/Last year's primary business income</i>	27	PAD/PAD Tahun Lalu**
34	<i>Account receivable /Accounts receivable of last year</i>	28	Piutang/Piutang Tahun Lalu
35	<i>Total assets/Total assets of last year</i>	29	Total Aset/Total Aset Tahun Lalu
36		30	Belanja Modal/Perubahan Nilai Aset Tetap ***
37		31	Belanja Barang Jasa/Perubahan Nilai Persediaan ***
38		32	Belanja Pegawai/Total Belanja ***



Tabel 3.2 Modifikasi Variabel Keuangan Sektor Swasta (Lanjutan)

No	Variabel Laporan Keuangan Sektor Swasta [22]	No	Modifikasi Variabel
39		33	Belanja Modal/Total Belanja ***
40		34	Belanja Hibah/Total Belanja***
41		35	Belanja Bantuan Sosial/Total Belanja***
42		36	PAD/Pendapatan Transfer***
		37	Zona Wilayah ***
		38	Jenis Wilayah Administratif ***
	Keterangan: * = variabel tidak dipakai ** = variabel dimodifikasi untuk sektor publik *** = tambahan variabel baru sesuai karakteristik LK sektor publik di Indonesia		

Data kuantitatif dari LHP LK sesuai Tabel 3.2 yang masih berupa pdf akan diinput ke Ms. Excel untuk diproses lebih lanjut. Hal ini dilakukan karena LHP LK tidak hanya berisi data angka-angka laporan keuangan namun ada data kualitatif berupa kalimat/kata yang menjelaskan angka-angka laporan keuangan yang tidak perlu diinputkan ke dalam Excel.

Variabel sektor swasta yang tidak dipakai adalah variabel yang terkait dengan laba atau sulit dicari variabel modifikasiannya di sektor publik. Variabel tambahan dari sektor publik adalah variabel yang terkait dengan belanja karena entitas sektor publik cenderung bersifat *cost center entity*. Variabel dengan nomor 1 s.d. 29 pada Kolom Modifikasi Variabel adalah variabel hasil modifikasian dari variabel sektor swasta, sedangkan variabel dengan nomor 30 s.d. 36 pada Kolom Modifikasi Variabel adalah variabel tambahan dari sektor publik yang dipengaruhi oleh belanja pemerintah, sedangkan sisanya yaitu variabel nomor 37 dan 38 merupakan variabel tambahan sesuai karakteristik wilayah di Indonesia.

Variabel inti (variabel nomor 1, 2, 3, 4, 5 dan 6) dinyatakan dalam jutaan rupiah. Variabel perbandingan (variabel nomor 7 s.d. 36) merupakan perbandingan variabel inti dengan variabel inti atau variabel laporan keuangan lainnya yang tidak menjadi fitur penelitian, variabel-variabel yang dibandingkan tersebut juga dinyatakan dalam jutaan rupiah. Variabel nomor 37 terkait dengan pembagian zona wilayah menurut BPK RI (nilai data 0 untuk zona barat dan nilai data 1 untuk zona timur). Variabel nomor 38 terkait dengan jenis wilayah administratif di Indonesia (nilai data 0 untuk provinsi, nilai data 0,5 untuk kabupaten, dan nilai data 1 untuk kota).

### **3.3 Normalisasi Data**

Tahap ini adalah tahap persiapan data agar dapat diolah lebih lanjut menggunakan metode penambahan data. Adapun detail langkah tahap ini adalah:

#### **3.3.1 Pembersihan Data**

Pembersihan data dilakukan untuk menghilangkan gangguan/*noise* yang menyebabkan data tidak konsisten atau data tidak relevan. Biasanya data yang diperoleh oleh seorang peneliti tidak sempurna seratus persen. Ada kondisi seperti data yang hilang atau tidak terisi, data yang tidak valid, data pencilan/*outlier* atau data salah input. Pembersihan data yang akan dilakukan dalam penelitian ini adalah untuk mengatasi data dengan nilai tak hingga/*infinity* akibat adanya pembagian suatu data dengan bilangan nol. Data tak hingga diganti dengan nilai maksimal untuk variabel/fitur dimana terdapat data tak hingga tersebut.

#### **3.3.2 Skema Normalisasi Data**

Penelitian ini menggunakan normalisasi data secara linier dan non linier. Perbandingan kinerja antara data mentah (tidak dinormalisasi/kode N0), normalisasi linier (kode N1 dan N2) dan normalisasi nonlinier (kode N3 dan N4) akan dilakukan untuk mencari model terbaik klasifikasi. Model terbaik tersebut kemudian digunakan pada data uji untuk pengukuran kinerja klasifikasi.

Skema normalisasi secara linier terdiri dari dua skema yaitu skema dengan kode N1 dan N2. Skema N1 dilakukan dengan merubah data mentah ke data baru

dengan skop 0 s.d. 1 menggunakan Persamaan (3.1). Adapun Skema N2 dilakukan dengan merubah data mentah ke data baru dengan skop -1 s.d. 1 menggunakan Persamaan (3.2) berikut:

$$\hat{x}_{ik} = \frac{x_{ik} - \min(x_k)}{\max(x_k) - \min(x_k)} \quad (3.1)$$

$$\hat{x}_{ik} = \frac{2x_{ik} - (\max(x_k) + \min(x_k))}{\max(x_k) - \min(x_k)} \quad (3.2)$$

Skema normalisasi nonlinier juga terdiri dari dua skema yaitu skema dengan kode N3 dan N4. Skema N3 dilakukan dengan merubah data mentah ke data baru menggunakan Persamaan (3.3). Adapun Skema N4 dilakukan dengan merubah data mentah ke data baru menggunakan Persamaan (3.4) berikut:

$$\hat{x}_{ik} = Z\ score(z) = \frac{x_{ik} - \bar{x}_k}{S_k} \quad (3.3)$$

$$\hat{x}_{ik} = \frac{1}{1 + e^{-z}} \quad (3.4)$$

Dimana:

$\hat{x}_{ik}$  = data baru pada baris ke-i dan kolom-k

$\bar{x}_k$  = rerata kolom-k

$S_k$  = simpangan baku kolom-k

$e$  = konstanta Euler, kira-kira setara 2,71828

$\min(x_k)$  = nilai minimum pada kolom-k

$\max(x_k)$  = nilai maksimum pada kolom-k

Ringkasan skema normalisasi untuk penelitian ini disajikan pada Tabel 3.3 berikut:

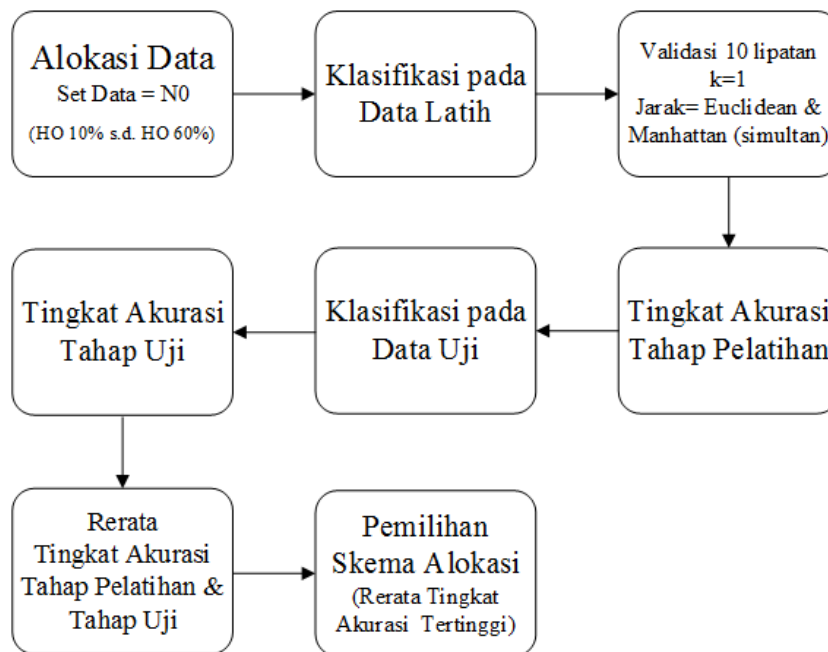
Tabel 3.3 Skema Normalisasi Data

N o	Kode Set Data	Uraian
1	N0	Data mentah (tanpa normalisasi).
2	N1	Normalisasi linier dengan skop (0,1) menggunakan Persamaan (3.1).
3	N2	Normalisasi linier dengan skop (-1,1) menggunakan Persamaan (3.2).
4	N3	Normalisasi nonlinier dengan <i>Z-score</i> menggunakan Persamaan (3.3).
5	N4	Normalisasi nonlinier dengan eksponensial <i>Z-score</i> menggunakan Persamaan (3.4).

### 3.4 Alokasi Data

Perbandingan antara metode *hold-out* (HO) 10% s.d. 60% dilakukan untuk mencari alokasi data latih dan data uji. Skema yang digunakan untuk pencarian model alokasi terbaik adalah validasi silang sepuluh lipatan/*10-folds cross validation* pada data latih menggunakan set data dengan kode N0 (data mentah asli tanpa normalisasi), dengan nilai-K yang dipakai dalam algoritma KNN adalah 1. Penentuan model alokasi data terbaik dilakukan dengan memilih nilai rerata akurasi tertinggi pada tahap pelatihan dan tahap uji pada setiap skema alokasi baik untuk klasifikasi 2 kelas maupun 4 kelas. Jarak *Euclidean* dan *Manhattan* digunakan secara simultan dalam pencarian model alokasi terbaik, nilai rerata akurasi klasifikasi menggunakan kedua jarak tersebut digunakan sebagai nilai skema alokasi (mulai dari skema HO 10% s.d. HO 60%).

Ringkasan skema pemilihan alokasi data terbaik untuk klasifikasi 2 kelas maupun 4 kelas diilustrasikan pada Gambar 3.2.



Gambar 3.2 Skema pemilihan alokasi data terbaik

### 3.5 Penggunaan Skema Fitur

Setelah diperoleh alokasi data latih dan data uji, baik untuk klasifikasi 2 kelas atau 4 kelas, maka dilanjutkan dengan pencarian skema fitur yang akan digunakan dalam proses selanjutnya. Kesuksesan suatu proses seleksi fitur menurut [45] adalah jika seleksi fitur tersebut mampu:

- Meningkatkan kinerja prediksi dari variabel prediktor;
- Menyediakan variabel prediktor yang cepat (*time efficiency*) dan efektif secara biaya (*cost-effective*);
- Memberikan pemahaman yang lebih baik dalam proses yang mendasari generalisasi data.

Rancangan skema fitur dalam penelitian ini digunakan untuk mencari model klasifikasi terbaik. Ada 9 skema fitur yang digunakan dalam penelitian ini. Skema nomor 1 s.d. 3 tidak menggunakan uji statistik untuk penentuan fiturnya, sedangkan skema nomor 4 s.d. 9 merupakan skema yang menggunakan uji statistik untuk penentuan fiturnya.

Skema 1 menggunakan fitur hasil modifikasian dari sektor swasta (fitur nomor 1 s.d. 29). Skema 2 menggunakan fitur tambahan dari sektor publik yang cenderung berorientasi pada belanja (fitur nomor 30 s.d. 38). Skema 3 menggunakan fitur gabungan antara Skema 1 dan Skema 2 (fitur nomor 1 s.d. 38).

Skema 4 menggunakan semua fitur yang signifikan sesuai Uji Wilcoxon Mann Whitney (2 kelas) dan Uji Kruskal-Wallis yang dilanjutkan dengan Uji Tukey *HSD* (4 kelas). Skema 5 menggunakan fitur dari Skema 4 yang memiliki nilai log<sub>10</sub> terkecil dari *p-value* dengan pembulatan hingga ke satuan terkecil (0 angka di belakang koma). Skema 6 menggunakan fitur dari Skema 4 yang memiliki signifikansi lebih dari 1 pengujian antar kelas pada klasifikasi 4 kelas.

Skema 7 menggunakan semua fitur yang signifikan sesuai Uji Wilcoxon Mann Whitney (2 kelas) dan Uji Kruskal-Wallis yang dilanjutkan dengan Uji Wilcoxon Mann Whitney (4 kelas). Skema 8 menggunakan fitur dari Skema 7 yang memiliki nilai log<sub>10</sub> terkecil dari *p-value* dengan pembulatan hingga ke satuan terkecil (0 angka di belakang koma). Skema 9 menggunakan fitur dari Skema 7 yang memiliki signifikansi lebih dari 1 pengujian antar kelas pada klasifikasi 4 kelas.

Simpulan skema fitur dalam penelitian ini disajikan pada Tabel 3.4 berikut ini:

Tabel 3.4 Skema Fitur Penelitian

No	Kode	Uraian
1	F1	<b>2 kelas dan 4 kelas:</b> Klasifikasi menggunakan fitur modifikasian dari sektor swasta ke sektor publik (fitur nomor 1 s.d. 29).
2	F2	<b>2 kelas dan 4 kelas:</b> Klasifikasi menggunakan fitur tambahan dari sektor publik yang cenderung bersifat belanja (fitur nomor 30 s.d. 38).
3	F3	<b>2 kelas dan 4 kelas:</b> Klasifikasi menggunakan seluruh fitur atau gabungan antara F1 dan F2 (fitur nomor 1 s.d. 38).
4	F4	<b>2 kelas:</b> Klasifikasi menggunakan fitur signifikan sesuai hasil Uji Wilcoxon Mann Whitney. <b>4 kelas:</b> Klasifikasi menggunakan fitur signifikan sesuai hasil Uji Kruskal-Wallis yang dilanjutkan dengan Uji Tukey <i>HSD</i> .

Tabel 3.4 Skema Fitur Penelitian (Lanjutan)

No	Kode	Uraian
5	F5	<b>2 kelas dan 4 kelas:</b> Klasifikasi menggunakan fitur pada F4 yang memiliki nilai log10 terkecil dari <i>p-value</i> dengan pembulatan hingga ke satuan terkecil (0 angka di belakang koma).
6	F6	<b>2 kelas dan 4 kelas:</b> Klasifikasi menggunakan fitur pada F4 yang memiliki signifikansi lebih dari 1 pengujian antar kelas.
7	F7	<b>2 kelas:</b> Klasifikasi menggunakan fitur signifikan sesuai hasil Uji Wilcoxon Mann Whitney. <b>4 kelas:</b> Klasifikasi menggunakan fitur signifikan sesuai hasil Uji Kruskal-Wallis yang dilanjutkan dengan Uji Wilcoxon Mann Whitney.
8	F8	<b>2 kelas dan 4 kelas:</b> Klasifikasi menggunakan fitur pada F7 yang memiliki nilai log10 terkecil dari <i>p-value</i> dengan pembulatan hingga ke satuan terkecil (0 angka di belakang koma).
9	F9	<b>2 kelas dan 4 kelas:</b> Klasifikasi menggunakan fitur pada F7 yang memiliki signifikansi lebih dari 1 pengujian antar kelas.

Metode Uji Wilcoxon Mann Whitney dan Uji Kruskal-Wallis dipilih karena termasuk uji nonparametrik, sehingga tidak memerlukan uji asumsi klasik seperti uji normalitas. Ketidakharian untuk melakukan uji asumsi klasik menjadikan uji nonparametrik lebih fleksibel digunakan dalam penelitian dibandingkan dengan uji parametrik.

### 3.6 Penggunaan Skema Nilai-K

Salah satu isu penting yang dalam metode KNN adalah pemilihan nilai-K [25]. Jika nilai-K terlalu kecil, maka hasil prediksi akan sensitif terhadap gangguan/*noise*. Di sisi lain, jika nilai-K terlalu besar, maka ada risiko bahwa data yang dipilih tidak relevan karena mayoritas data yang dipilih bisa dari data yang berlainan kelas atau terlalu jauh dari data yang ingin dikenali label kelasnya.

Penelitian ini menggunakan skema nilai-K sebagai berikut:

- Menggunakan nilai K=1 (kode 1-NN), dengan asumsi bahwa tetangga yang paling dekat adalah 1 tetangga.

- b. Menggunakan nilai K pada model klasifikasi terbaik (kode B-NN). Nilai-K pada skema ini didapatkan dari penggunaan metode validasi silang sepuluh lipatan/10-folds *crossvalidation*. Metode ini dilakukan bersama-sama dengan skema jarak (Euclidean dan Manhattan) pada data latih. Metode ini dilakukan terhadap masing-masing dataset (N0 s.d. N4) menggunakan nilai-K mulai 1 s.d. 90% jumlah data latih. Metode 10-folds *validation* digunakan untuk mencari model terbaik karena lebih kekar/*robust* dan akurat dibandingkan dengan metode lain seperti *Holdout*, *Leave-One-Out*, dan *Bootstrap* [46]. Untuk mencari model terbaik, proses ini dilakukan sebanyak sepuluh kali perulangan dengan menggunakan nilai rerata akurasi tertinggi sebagai model terbaik.

### 3.7 Penggunaan Skema Jarak

Skema jarak pada penelitian ini dibatasi hanya menggunakan Jarak Euclidean dan Manhattan. Kedua jarak ini dipilih karena merupakan jarak yang paling populer digunakan. Skema jarak ini dilakukan bersama-sama dengan skema nilai-K, skema fitur, dan skema normalisasi data untuk mencari model klasifikasi terbaik. Setelah semua tahap penggunaan berbagai skema penelitian dilakukan maka tahap selanjutnya adalah membandingkan kinerja tiap skema untuk mengetahui adakah perbedaan yang signifikan diantara masing-masing skema..

### 3.8 Perbandingan Kinerja Skema

Perbandingan skema fitur, skema normalisasi data, skema jarak, dan skema nilai-K dilakukan menggunakan Uji Friedman dan dilanjutkan dengan Uji Tukey *HSD*. Skema yang ingin diketahui perbedaan pengaruhnya akan diisolasi sebagai perlakuan penelitian dan kombinasi skema lain akan menjadi blok penelitian yang mendapatkan pengukuran berulang sesuai dengan skema perlakuan. Ilustrasi untuk pengujian pengaruh perlakuan skema jarak terhadap blok kombinasi skema lain menggunakan nilai rerata harmonik *F-Measure* dan AUC (RH) sebagai berikut:



Tabel 3.5 Ilustrasi Pengujian Perbedaan Pengaruh Skema Jarak

Blok	Blok Penelitian (Kombinasi Skema Lain)			Perlakuan Penelitian	
	Nilai-K	Normalisasi	Fitur	Euclidean (RH)	Manhattan (RH)
1	1-NN	N0	F1	0,4260	0,4502
2	1-NN	N0	F2	0,3284	0,3519
3	1-NN	N0	F3	0,4260	0,4502
4	1-NN	N0	F4	0,3902	0,4246

*Halaman ini sengaja dikosongkan*

## BAB 4

### HASIL DAN PEMBAHASAN

Bab ini akan menguraikan hasil dan pembahasan dari metode penelitian yang telah dilakukan pada tahap sebelumnya. Apakah klasifikasi data keuangan menggunakan fitur modifikasian dari sektor swasta dapat dijadikan metode alternatif untuk prediksi opini sektor publik adalah pertanyaan yang harus dijawab melalui metode dalam penelitian ini.

#### 4.1 Statistik Deskriptif Data Penelitian

Statistik Deskriptif dari data 150 LHP LK sesuai dengan fitur penelitian sebagaimana telah diuraikan pada Tabel 3.2, disajikan pada Tabel 4.1 berikut:

Tabel 4.1 Statistik Deskriptif Data Penelitian

No. Fitur	Jumlah Data	Nilai Minimal	Nilai Rerata	Nilai Maksimal	Simpangan Baku
1	150	0,00	80.437,06	2.373.399,00	282.212,08
2	150	10.361,00	5.347.201,57	102.266.081,00	9.890.640,42
3	150	-17.380,00	267.078,79	3.981.422,00	552.354,38
4	150	5.520,00	545.952,51	8.731.096,00	1.312.684,49
5	150	-17.380,00	263.811,35	3.985.017,00	548.002,87
6	150	0,00	84.126,75	2.304.198,00	234.711,53
7	150	-2,14	1,71	14,03	2,42
8	150	0,00	0,04	4,74	0,39
9	150	-0,02	0,24	28,04	2,29
10	150	-0,02	0,42	52,19	4,26
11	150	-2,14	1,62	13,89	2,23
12	150	0,00	0,31	3,88	0,48
13	150	0,00	1,06	149,17	12,17
14	150	-1,18	1.453,22	31.322,12*)	6.159,75
15	150	0,00	0,08	0,52	0,10
16	150	-0,02	0,24	28,60	2,33
17	150	-5,16	1.300,27	28.331,71*)	5.571,52
18	150	-3,24	0,00	0,86	0,28
19	150	0,00	0,00	0,04	0,01
20	150	0,00	0,02	1,45	0,12
21	150	-2,24	1,00	1,86	0,28
22	150	0,00	0,00	0,05	0,01
23	150	0,44	3,44	385,53	31,41

Tabel 4.1 Statistik Deskriptif Data Penelitian (Lanjutan)

No. Fitur	Jumlah Data	Nilai Minimal	Nilai Rerata	Nilai Maksimal	Simpangan Baku
24	150	-8,58	0,65	1,13	0,79
25	150	0,00	0,14	18,85	1,54
26	150	0,71	1,15	1,58	0,14
27	150	0,55	1,36	3,88	0,46
28	150	0,00	17,41	391,44*)	71,14
29	150	0,50	2,46	198,41	16,11
30	150	-65,93	0,42	10,83	5,68
31	150	-47.251,40	-456,40	5.169,70	4.590,21
32	150	0,09	0,48	2,13	0,21
33	150	0,05	0,24	2,79	0,24
34	150	0,00	0,06	1,14	0,11
35	150	0,00	0,02	0,48	0,04
36	150	0,01	0,31	4,08	0,58
37	150	0,00	0,45	1,00	0,50
38	150	0,00	0,54	1,00	0,34

Keterangan:

\*) = nilai maksimal hasil pembersihan data

#### 4.2 Hasil Proses Normalisasi Data

Salah satu proses normalisasi data dalam penelitian ini adalah pembersihan data pencilan dengan nilai tak hingga/*infinity*. Nilai tersebut diganti dengan nilai maksimal dimana terdapat data pencilan tersebut. Pembersihan data dilakukan untuk mempermudah proses klasifikasi selanjutnya. Rincian data yang mengalami proses pembersihan sebagai berikut:

Tabel 4.2 Data yang Mengalami Proses Pembersihan Data

No. Fitur	Fitur	Nomor Data	Jumlah Data	Nilai Sebelum Pembersihan	Nilai Sesudah Pembersihan
14	Aset Lancar/ Hutang Lancar	7, 57, 69, 71, 73	5	$\infty$ (tak hingga)	31.322,12
17	Kas Tunai dan Kas Bank/Hutang Lancar	7, 57, 69, 71, 73	5	$\infty$ (tak hingga)	28.331,71
28	Piutang/ Piutang Tahun Lalu	51, 69, 83, 96	4	$\infty$ (tak hingga)	28.331,71

### 4.3 Hasil Proses Alokasi Data

Proses ini dilakukan untuk mendapatkan skema alokasi data terbaik bagi klasifikasi 2 kelas dan 4 kelas. Nilai rerata laju akurasi tertinggi pada tahap pelatihan dan tahap pengujian menjadi kriteria untuk penentuan skema alokasi. Hasil pengujian skema alokasi data disajikan sebagai berikut:

#### 4.3.1 Alokasi Data untuk Klasifikasi Dua Kelas

Pengujian skema alokasi data untuk klasifikasi dua kelas menunjukkan bahwa Metode *Hold-Out* 30% (30% menjadi data uji dan 70% sisanya menjadi data latih) adalah skema alokasi terbaik dengan nilai rerata akurasi sebesar 52,94%. Rincian hasil pengujian skema alokasi data diurutkan dari rerata akurasi tertinggi sebagai berikut:

Tabel 4.3 Rerata Akurasi Proses Alokasi Data untuk Dua Kelas

Persentase <i>Hold-Out</i> (dalam satuan %)	Rerata Akurasi
30	0,5294
40	0,5289
50	0,5210
10	0,5081
60	0,5019
20	0,4581

#### 4.3.2 Alokasi Data untuk Klasifikasi Empat Kelas

Pengujian skema alokasi data untuk klasifikasi empat kelas menunjukkan bahwa Metode *Hold-Out* 60% (60% menjadi data uji dan 40% sisanya menjadi data latih) adalah skema alokasi terbaik dengan nilai rerata akurasi sebesar 45,25%. Rincian hasil pengujian skema alokasi data untuk klasifikasi empat kelas diurutkan dari rerata akurasi tertinggi sebagai berikut:

Tabel 4.4 Rerata Akurasi Proses Alokasi Data untuk Empat Kelas

Persentase <i>Hold-Out</i> (dalam satuan %)	Rerata Akurasi
60	0,4525
50	0,3960

Tabel 4.4 Rerata Akurasi Proses Alokasi Data untuk Empat Kelas (Lanjutan)

Persentase <i>Hold-Out</i> (dalam satuan %)	Rerata Akurasi
20	0,3950
10	0,3815
40	0,3778
30	0,3764

#### 4.4 Hasil Statistik Uji Signifikansi untuk Seleksi Fitur

Uji signifikansi dilakukan untuk memilih fitur yang signifikan secara statistik untuk digunakan dalam proses klasifikasi selanjutnya. Hipotesisnya adalah fitur yang signifikan secara statistik akan memberikan hasil klasifikasi yang lebih baik daripada fitur yang tidak signifikan.

##### 4.4.1 Seleksi Fitur untuk Klasifikasi Dua Kelas

Uji Wilcoxon Mann Whitney digunakan untuk penentuan signifikansi fitur pada klasifikasi dua kelas. Uji ini dilakukan untuk mengetahui adakah perbedaan signifikan antara kelas 1 (WTP) dan kelas 2 (selain WTP). Kedua grup termasuk dalam dua sampel yang independen (tidak ada satu sampel yang menjadi anggota di kedua grup tersebut), sehingga Uji ini dipilih sebagai penentu signifikansi fitur. Fitur signifikan pada klasifikasi 2 kelas ini akan sama pada F4 dan F7 serta F5 dan F8.

##### 4.4.1.1 Hasil Uji Wilcoxon Mann Whitney

Fitur-fitur laporan keuangan (fitur 1 s.d. fitur 38) diuji secara statistik untuk diketahui signifikansinya. Tingkat  $\alpha$  yang digunakan adalah 1%. Fitur yang memiliki Nilai-P < 1% adalah fitur yang dianggap signifikan secara statistik. Rincian fitur yang signifikan disajikan pada Tabel 4.5 berikut:

Tabel 4.5 Fitur Signifikan Sesuai Uji Wilcoxon Mann Whitney

Nomor Fitur	Nilai-P
2	0,0047
4	0,0000
6	0,0038

Tabel 4.5 Fitur Signifikan Sesuai Uji Wilcoxon Mann Whitney (Lanjutan)

Nomor Fitur	Nilai-P
7	0,0017
11	0,0010
13	0,0000
15	0,0000
24	0,0079
26	0,0015
35	0,0050
36	0,0000

#### 4.4.1.2 Nilai terkecil dari Log10 (Nilai-P) Uji Wilcoxon Mann Whitney

Semakin kecil Nilai-P suatu fitur, maka diasumsikan semakin kuat kemampuannya untuk membedakan antar kelas yang diklasifikasikan. Nilai Log10 dengan pembulatan hingga satuan terkecil dari fitur yang signifikan pada tahap sebelumnya dipilih sebagai fitur signifikan pada skema selanjutnya. Rincian nilai Log10 urut dari terkecil ke terbesar disajikan pada Tabel 4.6. Tanda bintang (\*) pada Fitur Nomor 36, 13, dan 15 menunjukkan bahwa mereka adalah fitur yang signifikan pada tahap ini.

Tabel 4.6 Nilai Log10 dari Nilai-P Uji Wilcoxon Mann Whitney

Nomor Fitur	Log10 (Nilai-P)
36*	-5
13*	-5
15*	-5
4	-4
11	-3
26	-3
7	-3
6	-2
2	-2
35	-2
24	-2

#### 4.4.1.3 Fitur yang memiliki signifikansi lebih dari 1 pengujian antar kelas

Fitur signifikan pada tahap ini akan mengikuti hasil uji dari klasifikasi 4 kelas, karena pada klasifikasi 2 kelas hanya ada pengujian signifikansi antar kelas

1 dan kelas 2. Skema Fitur F6 dan F9 baik untuk klasifikasi 2 kelas maupun 4 kelas akan menggunakan fitur signifikan dari hasil uji pada F6.

#### 4.4.2 Seleksi Fitur untuk Klasifikasi Empat Kelas

Uji Kruskal-Wallis digunakan untuk penentuan signifikansi fitur pada klasifikasi 4 kelas. Uji ini adalah uji signifikan yang bersifat umum untuk sampel independen lebih dari 2 grup. Uji lanjutan diperlukan untuk mengetahui antar grup manakah yang memiliki perbedaan signifikan. Penelitian ini menggunakan uji lanjutan Tukey *HSD* dan Uji Wilcoxon Mann Whitney. Kedua uji lanjutan ini juga akan dibandingkan untuk mengetahui mana yang memberikan kinerja klasifikasi lebih baik.

##### 4.4.2.1 Hasil Uji Kruskal-Wallis (KW) dilanjutkan Uji Tukey *HSD*

Fitur-fitur laporan keuangan (fitur 1 s.d. fitur 38) diuji secara statistik untuk diketahui signifikansinya. Tingkat  $\alpha$  yang digunakan adalah 1%. Fitur yang memiliki Nilai-P < 1% adalah fitur yang dianggap signifikan secara statistik. Rincian fitur yang signifikan disajikan pada Tabel 4.7 berikut:

Tabel 4.7 Fitur Signifikan Sesuai Uji KW dan Tukey *HSD*

Nomor Fitur	Nilai-P Kelas 1 x 3	Nilai-P Kelas 2 x 3
3	0,0003	0,0145
4	0,0008	0,0215
5	0,0004	0,0060
6	0,0075	0,0072
13	0,0017	0,0746
15	0,0026	0,0594
36	0,0016	0,0304

##### 4.4.2.2 Nilai terkecil dari Log10 (Nilai-P) sesuai Uji KW dan Tukey *HSD*

Semakin kecil Nilai-P suatu fitur, maka diasumsikan semakin kuat kemampuannya untuk membedakan antar kelas yang diklasifikasikan. Nilai Log10 dengan pembulatan hingga satuan terkecil dari fitur yang signifikan pada tahap sebelumnya dipilih sebagai fitur signifikan pada skema selanjutnya. Rincian nilai



Log10urut dari terkecil ke terbesar disajikan pada Tabel 4.8. Tanda bintang (\*) pada Fitur Nomor 3 menunjukkan bahwa hanya Fitur Nomor 3 yang menjadi fitur signifikan pada tahap ini.

Tabel 4.8 Nilai Log10 dari Nilai-P Uji KW dan Tukey *HSD*

Nomor Fitur	Kelas 1 x 3	Kelas 2 x 3
3*	-4	-2
5	-3	-2
4	-3	-2
36	-3	-2
13	-3	-1
15	-3	-1
6	-2	-2

#### 4.4.2.3 *Fitur yang memiliki signifikansi lebih dari 1 pengujian antar kelas sesuai Uji KW dan Tukey HSD*

Asumsi yang mendasari pemilihan fitur ini adalah suatu fitur yang memiliki perbedaan signifikan lebih dari 1 pengujian antar grup akan memberikan hasil klasifikasi lebih baik dari fitur yang hanya memiliki perbedaan signifikan dalam 1 pengujian saja. Fitur Nomor 5 dan 6 merupakan fitur yang signifikan pada tahap ini sebagaimana disajikan pada Tabel 4.9.

Tabel 4.9 Fitur Signifikan lebih dari satu uji Sesuai Uji KW-Tukey *HSD*

Nomor Fitur	Nilai-P Kelas 1 x 3	Nilai-P Kelas 2 x 3
5	0,0004	0,0060
6	0,0075	0,0072

#### 4.4.2.4 *Hasil Uji Kruskal-Wallis Dilanjutkan Uji Wilcoxon Mann Whitney*

Fitur-fitur laporan keuangan (fitur 1 s.d. fitur 38) diuji secara statistik untuk diketahui signifikansinya. Tingkat  $\alpha$  yang digunakan adalah 1%. Fitur yang memiliki Nilai-P  $< 1\%$  adalah fitur yang dianggap signifikan secara statistik menggunakan Uji Kruskal-Wallis yang dilanjutkan dengan Uji Wilcoxon Mann

Whitney. Rincian fitur yang signifikan pada tahap ini disajikan pada Tabel 4.10 berikut:

Tabel 4.10 Fitur Signifikan Sesuai Uji KW dan Uji WMW

Nomor Fitur	Nilai-P Kelas 1 x 3	Nilai-P Kelas 2 x 3	Nilai-P Kelas 3 x 4
2	0,0142	0,0073	0,1859
3	0,0002	0,0002	0,0046
4	0,0004	0,0022	0,0257
5	0,0002	0,0004	0,0046
6	0,0017	0,0013	0,0756
9	0,0072	0,0516	0,0359
10	0,0021	0,0245	0,0224
13	0,0012	0,0038	0,0448
15	0,0020	0,0028	0,0329
16	0,0079	0,0477	0,0358
36	0,0007	0,0021	0,2361

#### 4.4.2.5 Nilai Terkecil dari Log10 (Nilai-P) Sesuai Uji Kruskal-Wallis dan Uji

##### Wilcoxon Mann Whitney

Nilai Log10 dari Nilai-P terkecil pada tahap ini yaitu -4. Fitur yang signifikan pada tahap ini yaitu Fitur Nomor 3 dan 5. Rincian nilai Log10 dari Nilai-P pada tahap ini disajikan pada Tabel 4.11 berikut:

Tabel 4.11 Nilai Log10 dari Nilai-P Uji KW dan Uji WMW

Nomor Fitur	Kelas 1 x 3	Kelas 2 x 3	Kelas 3 x 4
5*	-4*	-3	-2
3*	-4*	-4*	-2
4	-3	-3	-2
36	-3	-3	-1
13	-3	-2	-1
6	-3	-3	-1
15	-3	-3	-1
10	-3	-2	-2
9	-2	-1	-1
16	-2	-1	-1
2	-2	-2	-1

#### 4.4.2.6 Fitur yang memiliki signifikansi lebih dari 1 pengujian antar kelas sesuai Uji Kruskal-Wallis dan Uji Wilcoxon Mann Whitney

Tahap ini memberikan fitur yang lebih banyak dibandingkan tahap pada Uji Kruskal-Wallis dan Uji Tukey *HSD*. Fitur Nomor 3, 4, 5, 6, 13, 15, dan 36 menjadi fitur signifikan pada tahap ini. Fitur Nomor 3 dan 5 bahkan memiliki signifikansi pada 3 pengujian antar kelas. Rincian hasil uji tahap ini disajikan pada Tabel 4.12 berikut:

Tabel 4.12 Fitur Signifikan lebih dari satu uji Sesuai Uji KW-WMW

Nomor Fitur	Nilai-P Kelas 1 x 3	Nilai-P Kelas 2 x 3	Nilai-P Kelas 3 x 4
3	0,0002*	0,0002*	0,0046*
4	0,0004*	0,0022*	0,0257*
5	0,0002*	0,0004*	0,0046*
6	0,0017*	0,0013*	0,0756
13	0,0012*	0,0038*	0,0448*
15	0,0020*	0,0028*	0,0329*
36	0,0007*	0,0021*	0,2361

#### 4.5 Implementasi Hasil Uji Statistik pada Skema Fitur

Skema fitur yang digunakan dalam penelitian ini terbagi menjadi dua bagian. Fitur F1 s.d. F3 tidak melalui uji statistik, sedangkan Fitur F4 s.d. F9 adalah fitur yang menggunakan uji statistik dalam penentuannya. Detail skema fitur setelah tahap uji statistik disajikan pada Tabel 4.13 berikut:

Tabel 4.13 Skema Fitur Penelitian Lengkap Setelah Uji Statistik

No	Kode	Uraian
1	F1	2 kelas dan 4 kelas: fitur nomor 1 s.d. 29.
2	F2	2 kelas dan 4 kelas: fitur nomor 30 s.d. 38.
3	F3	2 kelas dan 4 kelas: fitur nomor 1 s.d. 38.
4	F4	2 kelas: fitur nomor 2, 4, 6, 7, 11, 13, 15, 24, 26, 35, 36. 4 kelas: fitur nomor 3, 4, 5, 6, 13, 15, 36.

Tabel 4.13 Skema Fitur Penelitian Lengkap Setelah Uji Statistik (Lanjutan)

No	Kode	Uraian
5	F5	2 kelas: fitur nomor 36, 13, 15. 4 kelas: fitur nomor 3.
6	F6	2 kelas dan 4 kelas: fitur nomor 5, 6.
7	F7	2 kelas: <b>sama dengan fitur pada F4.</b> 4 kelas: fitur nomor 2, 3, 4, 5, 6, 9, 10, 13, 15, 16, 36.
8	F8	2 kelas: <b>sama dengan fitur pada F5.</b> 4 kelas: fitur nomor 3, 5.
9	F9	2 kelas dan 4 kelas: fitur nomor 3, 4, 5, 6, 13, 15, 36.

#### 4.6 Sepuluh Model Terbaik Berdasarkan Rerata Harmonik *F-Measure* dan AUC (RH)

Hasil klasifikasi menggunakan berbagai skema normalisasi, fitur, nilai-K, dan jarak menggunakan data uji dijadikan sebagai media evaluasi kinerja klasifikasi. Kinerja klasifikasi dipisahkan untuk klasifikasi 2 kelas dan 4 kelas. Rincian pengukuran kinerja klasifikasi tersebut sebagai berikut:

##### 4.6.1 Sepuluh Model Terbaik Klasifikasi 2 Kelas

Model dengan kombinasi N4, F4 atau F7, Euclidean, B-NN dengan K=30 menghasilkan RH tertinggi sebesar 0,6601. Skema-skema yang unggul dari skema lainnya jika dilihat dari 10 model terbaik adalah:

- N1 dan N2 dengan frekuensi kemunculan masing-masing 3 kali.
- Jarak Manhattan dengan frekuensi kemunculan 6 kali.
- Skema 1-NN dengan frekuensi kemunculan 8 kali.
- Skema seleksi fitur F4 atau F7 dengan frekuensi kemunculan 4 kali.

Adapun rincian dari 10 model terbaik dari 140 kombinasi skema pada klasifikasi 2 kelas disajikan pada Tabel 4.14 berikut:

Tabel 4.14 Sepuluh Model Terbaik Klasifikasi 2 Kelas

No	Normalisasi Data	Fitur	Jarak	Nilai-K	RH
1	N4	F4 atau F7	Euclidean	B-NN (30)	0,6601
2	N4	F5 atau F8	Euclidean	1-NN	0,6443
3	N1	F3	Euclidean	1-NN	0,6433
4	N2	F3	Euclidean	1-NN	0,6433
5	N3	F4 atau F7	Manhattan	B-NN (23)	0,6315
6	N1	F4 atau F7	Manhattan	1-NN	0,6224
7	N2	F4 atau F7	Manhattan	1-NN	0,6224
8	N1	F2	Manhattan	1-NN	0,6223
9	N2	F2	Manhattan	1-NN	0,6223
10	N0	F5 atau F8	Manhattan	1-NN	0,6223

#### 4.6.2 Sepuluh Model Terbaik Klasifikasi 4 Kelas

Model dengan kombinasi N1, F2, Manhattan, B-NN dengan K=3 menghasilkan RH tertinggi sebesar 0,5378. Skema-skema yang unggul dari skema lainnya jika dilihat dari 10 model terbaik adalah:

- N4 dengan frekuensi kemunculan 4 kali.
- Jarak Euclidean dengan frekuensi kemunculan 6 kali.
- Skema B-NN dengan frekuensi kemunculan 6 kali.
- F2 dengan frekuensi kemunculan 10 kali.

Adapun rincian dari 10 model terbaik dari 180 kombinasi skema pada klasifikasi 4 kelas disajikan pada Tabel 4.15 berikut:

Tabel 4.15 Sepuluh Model Terbaik Klasifikasi 4 Kelas

No	Normalisasi Data	Fitur	Jarak	Nilai-K	RH
1	N1	F2	Manhattan	B-NN (3)	0,5378
2	N1	F2	Euclidean	1-NN	0,5352
3	N2	F2	Euclidean	1-NN	0,5352
4	N1	F2	Euclidean	B-NN (2)	0,5352
5	N2	F2	Euclidean	B-NN (2)	0,5352
6	N3	F2	Manhattan	B-NN (3)	0,5346
7	N4	F2	Manhattan	1-NN	0,5341
8	N4	F2	Manhattan	B-NN (3)	0,5309
9	N4	F2	Euclidean	1-NN	0,5296
10	N4	F2	Euclidean	B-NN (2)	0,5296

#### 4.7 Hasil Uji Friedman pada Skema Normalisasi Data

Uji ini dilakukan untuk mengetahui apakah ada perbedaan signifikan antara penggunaan satu skema dengan skema lainnya untuk lebih dari dua kelompok skema. Uji Friedman pada skema normalisasi data (SND) berarti pengujian adakah perbedaan yang signifikan pada hasil klasifikasi dengan menggunakan skema N0, N1, N2, N3 dan N4. Hasil pengujian skema dengan hasil terbaik direkomendasikan untuk digunakan dalam penelitian selanjutnya.

##### 4.7.1 Hasil Uji SND pada Klasifikasi 2 Kelas

Hasil rerata peringkat menunjukkan bahwa skema normalisasi data N3 memiliki rerata peringkat tertinggi sebesar 3,33 disusul skema N4 dengan rerata peringkat 3,31. Jika menggunakan hasil pemeringkatan skema normalisasi data, maka direkomendasikan penggunaan skema N3 dan N4 (skema nonlinier), karena skema tersebut mengungguli skema lain. Rincian urutan skema dengan rerata peringkat tertinggi ke rerata peringkat terendah disajikan pada Tabel 4.16, sedangkan rincian proses pengujian pengaruh Skema Normalisasi Data terhadap blok uji diuraikan pada Lampiran 1.

Tabel 4.16 Rerata dan Varian Peringkat SND Klasifikasi 2 Kelas

No	Normalisasi Data	Rerata Peringkat	Varian Peringkat
1	N3	3,33	1,11
2	N4	3,31	1,26
3	N2	3,28	1,37
4	N1	2,90	1,30
5	N0	2,18	1,46

Hasil Uji Friedman atas skema normalisasi data pada tingkat taraf nyata ( $\alpha=5\%$ ) menunjukkan bahwa tidak ada perbedaan signifikan antara penggunaan 4 Skema dengan rerata peringkat tertinggi (N3, N4, N2, dan N1 dengan Nilai-P untuk pengujian antar skema tersebut memiliki nilai  $> 5\%$ , yaitu antara 75% s.d. 100%). Uji Friedman juga menunjukkan tidak ada perbedaan signifikan pada penggunaan 2 skema dengan rerata peringkat terendah (N0 dan N1 dengan Nilai-P = 25,7%). Akan tetapi, kinerja Skema N0 sangat buruk jika dibandingkan dengan 3 skema

dengan rerata peringkat tertinggi (N3, N4, dan N2) dengan Nilai-P antara 1,1% s.d. 1,8%. Berdasarkan hasil Uji Friedman tersebut direkomendasikan untuk tidak menggunakan Skema N0 (data asli tanpa normalisasi), sebaliknya direkomendasikan untuk menggunakan skema normalisasi linier (N2) atau skema normalisasi nonlinier (N3 dan N4) dalam klasifikasi 2 kelas.

Rincian hasil Uji Friedman untuk skema normalisasi data pada klasifikasi 2 kelas disajikan pada Tabel 4.17 (tanda \* pada Kolom Nilai-P menunjukkan ada perbedaan signifikan antar grup yang dibandingkan).

Tabel 4.17 Hasil Uji Friedman pada SND Klasifikasi 2 Kelas

No	Grup A	Grup B	Rerata Peringkat Grup A	Rerata Peringkat Grup B	Selisih	Nilai-P
1	N0	N1	2,18	2,90	0,72	0,257
2	N0	N2	2,18	3,28	1,10	0,018*
3	N0	N3	2,18	3,33	1,15	0,011*
4	N0	N4	2,18	3,31	1,13	0,014*
5	N1	N2	2,90	3,28	0,38	0,833
6	N1	N3	2,90	3,33	0,43	0,750
7	N1	N4	2,90	3,31	0,40	0,793
8	N2	N3	3,28	3,33	0,06	1,000
9	N2	N4	3,28	3,31	0,03	1,000
10	N3	N4	3,33	3,31	0,03	1,000

#### 4.7.2 Hasil Uji SND pada Klasifikasi 4 Kelas

Hasil rerata peringkat menunjukkan bahwa skema normalisasi data N4 memiliki rerata peringkat tertinggi sebesar 3,47 disusul skema N3 dengan rerata peringkat 3,31. Jika menggunakan hasil pemeringkatan skema normalisasi data, maka direkomendasikan penggunaan skema N4 dan N3 (skema nonlinier), karena skema tersebut mengungguli skema lain. Rincian urutan skema dengan rerata peringkat tertinggi ke rerata peringkat terendah disajikan pada Tabel 4.18, sedangkan rincian proses pengujian pengaruh Skema Normalisasi Data terhadap blok uji diuraikan pada Lampiran 2.

Tabel 4.18 Rerata dan Varian Peringkat SND Klasifikasi 4 Kelas

No	Normalisasi Data	Rerata Peringkat	Varian Peringkat
1	N4	3,47	1,36
2	N3	3,31	1,17
3	N1	3,07	1,19
4	N2	2,76	1,08
5	N0	2,39	1,51

Hasil Uji Friedman atas skema normalisasi data pada tingkat taraf nyata ( $\alpha=5\%$ ) menunjukkan bahwa tidak ada perbedaan signifikan antara penggunaan 4 skema dengan rerata peringkat tertinggi (N4, N3, N1, dan N2 dengan Nilai-P untuk pengujian antar skema tersebut memiliki nilai  $> 5\%$ , yaitu antara 24,2% s.d. 98,9%). Uji Friedman juga menunjukkan tidak ada perbedaan signifikan pada penggunaan 4 skema terbawah (N0, N2, N1, dan N3 dengan Nilai-P antara 6,1% s.d. 81,4%). Akan tetapi, kinerja Skema N0 sangat buruk jika dibandingkan dengan skema N4 (Nilai-P 1,5%). Berdasarkan hasil Uji Friedman tersebut direkomendasikan untuk tidak menggunakan Skema N0 (data asli tanpa normalisasi), sebaliknya direkomendasikan untuk menggunakan skema normalisasi nonlinier N4 dalam klasifikasi 4 kelas.

Rincian hasil Uji Friedman untuk skema normalisasi data pada klasifikasi 4 kelas disajikan pada Tabel 4.19 (tanda \* pada Kolom Nilai-P menunjukkan ada perbedaan signifikan antar grup yang ditandingkan).

Tabel 4.19 Hasil Uji Friedman pada SND Klasifikasi 4 Kelas

No	Grup A	Grup B	Rerata Peringkat Grup A	Rerata Peringkat Grup B	Selisih	Nilai-P
1	N0	N1	2,39	3,07	0,68	0,281
2	N0	N2	2,39	2,76	0,38	0,814
3	N0	N3	2,39	3,31	0,92	0,061
4	N0	N4	2,39	3,47	1,08	0,015*
5	N1	N2	3,07	2,76	0,31	0,903
6	N1	N3	3,07	3,31	0,24	0,960
7	N1	N4	3,07	3,47	0,40	0,771
8	N2	N3	2,76	3,31	0,54	0,518
9	N2	N4	2,76	3,47	0,71	0,242
10	N3	N4	3,31	3,47	0,17	0,989



#### 4.8 Hasil Uji Friedman pada Skema Fitur

Uji ini dilakukan untuk mengetahui apakah ada perbedaan signifikan antara penggunaan satu skema dengan skema lainnya untuk lebih dari dua kelompok skema. Uji Friedman pada skema fitur (SF) berarti pengujian adakah perbedaan yang signifikan pada hasil klasifikasi dengan menggunakan skema F1 s.d. F9. Hasil pengujian skema dengan hasil terbaik direkomendasikan untuk digunakan dalam penelitian selanjutnya.

##### 4.8.1 Uji Friedman pada Skema Fitur Klasifikasi 2 Kelas

Hasil rerata peringkat menunjukkan bahwa skema fitur F5 atau F8 memiliki rerata peringkat tertinggi sebesar 6,55 disusul skema F3 dengan rerata peringkat 5,85. Pada klasifikasi 2 kelas, pasangan skema F4 dan F7 serta F5 dan F8 adalah skema fitur yang sama, kedua pasangan tersebut hanya akan berbeda pada klasifikasi 4 kelas. Jika menggunakan hasil pemeringkatan skema fitur, maka direkomendasikan penggunaan skema F5 atau F8 (Log10 terkecil dari fitur signifikan menurut Uji Wilcoxon Mann Whitney), karena skema tersebut mengungguli skema lain. Rincian urutan skema dengan rerata peringkat tertinggi ke rerata peringkat terendah disajikan pada Tabel 4.20, sedangkan rincian proses pengujian pengaruh Skema Fitur terhadap blok uji diuraikan pada Lampiran 3.

Tabel 4.20 Rerata dan Varian Peringkat Skema Fitur (2 Kelas)

No	Skema Fitur	Rerata Peringkat	Varian Peringkat
1	F5 atau F8	6,55	2,21
2	F3	5,85	2,06
3	F4 atau F7	5,78	1,92
4	F2	5,45	2,72
5	F1	3,70	2,18
6	F9	3,10	2,15
7	F6	2,25	1,59

Hasil Uji Friedman atas skema fitur pada tingkat taraf nyata ( $\alpha=5\%$ ) menunjukkan bahwa tidak ada perbedaan signifikan antara penggunaan 4 skema dengan rerata peringkat tertinggi (F5 atau F8, F3, F4 atau F7, dan F2). Uji Friedman juga menunjukkan kinerja Skema F1 lebih buruk secara signifikan dibandingkan

F5 atau F8. Skema F6 dan F9 memiliki kinerja sangat buruk jika dibandingkan dengan tiga skema dengan rerata peringkat tertinggi. Berdasarkan hasil Uji Friedman tersebut direkomendasikan untuk tidak menggunakan Skema F1, F6, dan F9, sebaliknya direkomendasikan untuk menggunakan Skema F5 atau F8 (Log10 terkecil dari fitur signifikan menurut Uji Wilcoxon Mann Whitney) dalam klasifikasi 2 kelas.

Rincian hasil Uji Friedman untuk skema fitur pada klasifikasi 2 kelas yang memiliki perbedaan signifikan disajikan pada Tabel 4.21 (tanda \* pada Kolom Nilai-P menunjukkan ada perbedaan signifikan antar grup yang ditandingkan).

Tabel 4.21 Perbedaan Signifikan pada Skema Fitur Klasifikasi 2 Kelas

No	Grup A	Grup B	Rerata Peringkat Grup A	Rerata Peringkat Grup B	Selisih	Nilai-P
1	F1	F5 atau F8	3,70	6,55	2,85	0,0242*
2	F2	F6	5,45	2,25	3,20	0,0057*
3	F3	F6	5,85	2,25	3,60	0,0009*
4	F3	F9	5,85	3,10	2,75	0,0352*
5	F4 atau F7	F6	5,78	2,25	3,53	0,0012*
6	F4 atau F7	F9	5,78	3,10	2,68	0,0461*
7	F5 atau F8	F6	6,55	2,25	4,30	0,0000*
8	F5 atau F8	F9	6,55	3,10	3,45	0,0018*

#### 4.8.2 Uji Friedman pada Skema Fitur Klasifikasi 4 Kelas

Hasil rerata peringkat menunjukkan bahwa skema fitur F2 memiliki rerata peringkat tertinggi sebesar 7,45 disusul skema F7 dengan rerata peringkat 6,53. Jika menggunakan hasil pemeringkatan skema normalisasi data, maka direkomendasikan penggunaan skema F2, karena skema tersebut mengungguli skema lain. Rincian urutan skema dengan rerata peringkat tertinggi ke rerata peringkat terendah disajikan pada Tabel 4.22, sedangkan rincian proses pengujian pengaruh Skema Fitur terhadap blok uji diuraikan pada Lampiran 4.

Tabel 4.22 Rerata dan Varian Peringkat Skema Fitur (4 Kelas)

No	Skema Fitur	Rerata Peringkat	Varian Peringkat
1	F2	7,45	3,19
2	F7	6,53	1,77
3	F4	5,95	1,32
4	F9	5,93	1,37
5	F3	5,80	2,30
6	F1	4,65	2,15
7	F8	4,60	1,85
8	F5	2,60	0,88
9	F6	1,50	1,00

Hasil Uji Friedman atas skema fitur pada tingkat taraf nyata ( $\alpha=5\%$ ) menunjukkan bahwa tidak ada perbedaan signifikan antara penggunaan 5 skema dengan rerata peringkat tertinggi (F2, F7, F4, F9, dan F3). Uji Friedman juga menunjukkan bahwa 4 skema dengan rerata peringkat terendah (F1, F8, F5, dan F6) memiliki kinerja yang lebih rendah secara signifikan jika dibandingkan dengan Skema F2. Berdasarkan hasil Uji Friedman tersebut direkomendasikan untuk tidak menggunakan 4 skema dengan rerata peringkat terendah (F1, F8, F5, dan F6), sebaliknya direkomendasikan untuk menggunakan Skema F2 (fitur asli dari sektor publik yang terkait dengan belanja pemerintah) yang memiliki rerata peringkat tertinggi untuk klasifikasi 4 kelas.

Rincian hasil Uji Friedman untuk skema fitur pada klasifikasi 4 kelas disajikan pada Tabel 4.23 (tanda \* pada Kolom Nilai-P menunjukkan ada perbedaan signifikan antar grup yang ditandingkan).

Tabel 4.23 Perbedaan Signifikan pada Skema Fitur Klasifikasi 4 Kelas

No	Grup A	Grup B	Rerata Peringkat Grup A	Rerata Peringkat Grup B	Selisih	Nilai-P
1	F1	F2	4,65	7,45	2,80	0,0318*
2	F1	F6	4,65	1,50	3,15	0,0079*
3	F2	F5	7,45	2,60	4,85	0,0000*
4	F2	F6	7,45	1,50	5,95	0,0000*
5	F2	F8	7,45	4,60	2,85	0,0264*
6	F3	F5	5,80	2,60	3,20	0,0064*
7	F3	F6	5,80	1,50	4,30	0,0000*

Tabel 4.23 Perbedaan Signifikan pada Skema Fitur Klasifikasi 4 Kelas - Lanjutan

No	Grup A	Grup B	Rerata Peringkat Grup A	Rerata Peringkat Grup B	Selisih	Nilai-P
8	F4	F5	5,95	2,60	3,35	0,0032*
9	F4	F6	5,95	1,50	4,45	0,0000*
10	F5	F7	2,60	6,53	3,93	0,0002*
11	F5	F9	2,60	5,93	3,33	0,0036*
12	F6	F7	1,50	6,53	5,03	0,0000*
13	F6	F8	1,50	4,60	3,10	0,0097*
14	F6	F9	1,50	5,93	4,43	0,0000*

#### 4.9 Hasil Uji Wilcoxon Peringkat Bertanda pada Skema Nilai-K

Uji ini dilakukan untuk mengetahui apakah ada perbedaan signifikan antara penggunaan satu skema dengan skema lainnya untuk dua kelompok skema. Uji Wilcoxon Peringkat Bertanda (WSR) pada skema Nilai-K berarti pengujian adakah perbedaan yang signifikan pada hasil klasifikasi dengan menggunakan Skema 1-NN dan B-NN. Hasil pengujian skema dengan hasil terbaik direkomendasikan untuk digunakan dalam penelitian selanjutnya.

##### 4.9.1 Uji WSR pada Skema Nilai-K Klasifikasi 2 Kelas

Hasil rerata peringkat menunjukkan bahwa Skema 1-NN memiliki rerata peringkat tertinggi sebesar 1,54 dibanding Skema B-NN dengan rerata peringkat 1,46. Rerata peringkat dan varian peringkat dari masing-masing skema disajikan pada Tabel 4.24, sedangkan rincian proses pengujian pengaruh Skema Nilai-K terhadap blok uji diuraikan pada Lampiran 5.

Tabel 4.24 Rerata dan Varian Peringkat Skema Nilai-K (2 Kelas)

No	Skema Nilai-K	Rerata Peringkat	Varian Peringkat
1	1-NN	1,54	0,45
2	B-NN	1,46	0,45

Hasil Uji Wilcoxon Peringkat Bertanda atas skema Nilai-K pada tingkat taraf nyata ( $\alpha=5\%$ ) menunjukkan bahwa tidak ada perbedaan signifikan antara penggunaan Skema 1-NN dan B-NN. Berdasarkan hasil Uji Wilcoxon Peringkat

Bertanda tersebut direkomendasikan penggunaan skema 1-NN untuk klasifikasi dengan 2 kelas label karena kemudahan proses, efisiensi waktu, dan akurasi hasil.

Rincian hasil Uji Wilcoxon Peringkat Bertanda untuk skema Nilai-K pada klasifikasi 2 kelas disajikan pada Tabel 4.25.

Tabel 4.25 Hasil Uji WSR pada Skema Nilai-K Klasifikasi 2 Kelas

No	Grup A	Grup B	Rerata Peringkat Grup A	Rerata Peringkat Grup B	Selisih	Nilai-P
1	1-NN	B-NN	1,54	1,46	0,08	0,1241

#### 4.9.2 Uji WSR pada Skema Nilai-K Klasifikasi 4 Kelas

Hasil rerata peringkat menunjukkan bahwa skema B-NN memiliki rerata peringkat tertinggi sebesar 1,64 dibanding skema 1-NN dengan rerata peringkat 1,36. Rerata peringkat dan varian peringkat dari masing-masing skema disajikan pada Tabel 4.26, sedangkan rincian proses pengujian pengaruh Skema Nilai-K terhadap blok uji diuraikan pada Lampiran 6.

Tabel 4.26 Rerata dan Varian Peringkat Skema Nilai-K (4 Kelas)

No	Skema Nilai-K	Rerata Peringkat	Varian Peringkat
1	B-NN	1,64	0,44
2	1-NN	1,36	0,44

Hasil Uji Wilcoxon Peringkat Bertanda atas skema Nilai-K pada tingkat taraf nyata ( $\alpha=5\%$ ) menunjukkan bahwa ada perbedaan signifikan antara penggunaan Skema B-NN dan 1-NN dengan Nilai-P sebesar 0,12% untuk keunggulan Skema B-NN. Berdasarkan hasil Uji Wilcoxon Peringkat Bertanda tersebut direkomendasikan penggunaan Skema B-NN untuk klasifikasi dengan 4 kelas label.

Rincian hasil Uji Wilcoxon Peringkat Bertanda untuk skema Nilai-K pada klasifikasi 4 kelas disajikan pada Tabel 4.27 (tanda \* pada Kolom Nilai-P menunjukkan ada perbedaan signifikan antar grup yang dibandingkan).

Tabel 4.27 Hasil Uji WSR pada Skema Nilai-K Klasifikasi 4 Kelas

No	Grup A	Grup B	Rerata Peringkat Grup A	Rerata Peringkat Grup B	Selisih	Nilai-P
1	1-NN	B-NN	1,36	1,64	0,28	0,0012*

#### 4.10 Hasil Uji Wilcoxon Peringkat Bertanda pada Skema Jarak

Uji ini dilakukan untuk mengetahui apakah ada perbedaan signifikan antara penggunaan satu skema dengan skema lainnya untuk dua kelompok skema. Uji Wilcoxon Peringkat Bertanda pada skema jarak berarti pengujian adakah perbedaan yang signifikan pada hasil klasifikasi dengan menggunakan jarak Euclidean dan Manhattan. Hasil pengujian skema dengan hasil terbaik direkomendasikan untuk digunakan dalam penelitian selanjutnya.

##### 4.10.1 Uji WSR pada Skema Jarak Klasifikasi 2 Kelas

Hasil rerata peringkat menunjukkan bahwa Skema Euclidean dan Manhattan memiliki rerata peringkat yang sama sebesar 1,50. Rerata peringkat dan varian peringkat dari masing-masing skema disajikan pada Tabel 4.28, sedangkan rincian proses pengujian pengaruh Skema Jarak terhadap blok uji diuraikan pada Lampiran 7.

Tabel 4.28 Rerata dan Varian Peringkat Skema Jarak (2 Kelas)

No	Skema Jarak	Rerata Peringkat	Varian Peringkat
1	Euclidean	1,50	0,47
2	Manhattan	1,50	0,47

Hasil Uji Wilcoxon Peringkat Bertanda atas skema jarak pada tingkat taraf nyata ( $\alpha=5\%$ ) menunjukkan bahwa tidak ada perbedaan signifikan antara penggunaan Skema Jarak Euclidean dan Manhattan (Nilai-P =15,57%). Berdasarkan hasil Uji Wilcoxon Peringkat Bertanda dan rerata peringkat masing-masing skema tidak ada rekomendasi yang dapat diberikan untuk klasifikasi dengan 2 kelas label.

Rincian hasil Uji Wilcoxon Peringkat Bertanda untuk skema jarak pada klasifikasi 2 kelas disajikan pada Tabel 4.29.

Tabel 4.29 Hasil Uji WSR pada Skema Jarak Klasifikasi 2 Kelas

No	Grup A	Grup B	Rerata Peringkat Grup A	Rerata Peringkat Grup B	Selisih	Nilai-P
1	Euclidean	Manhattan	1,50	1,50	0	0,1557

#### 4.10.2 Uji WSR pada Skema Jarak Klasifikasi 4 Kelas

Hasil rerata peringkat menunjukkan bahwa Skema Euclidean memiliki rerata peringkat tertinggi sebesar 1,54 dibanding skema Manhattan dengan rerata peringkat 1,46. Rerata peringkat dan varian peringkat dari masing-masing skema disajikan pada Tabel 4.30, sedangkan rincian proses pengujian pengaruh Skema Jarak terhadap blok uji diuraikan pada Lampiran 8.

Tabel 4.30 Rerata dan Varian Peringkat Skema Jarak (4 Kelas)

No	Skema Jarak	Rerata Peringkat	Varian Peringkat
1	Euclidean	1,54	0,47
2	Manhattan	1,46	0,47

Hasil Uji Wilcoxon Peringkat Bertanda atas skema jarak pada tingkat taraf nyata ( $\alpha=5\%$ ) menunjukkan bahwa tidak ada perbedaan signifikan antara penggunaan Skema Jarak Euclidean dan Manhattan (Nilai-P =33,44%). Berdasarkan hasil Uji Wilcoxon Peringkat Bertanda dan rerata peringkat masing-masing skema, Skema Euclidean direkomendasikan karena memiliki akurasi yang lebih tinggi dari Skema Manhattan.

Rincian hasil Uji Wilcoxon Peringkat Bertanda untuk skema jarak pada klasifikasi 4 kelas disajikan pada Tabel 4.31.

Tabel 4.31 Hasil Uji WSR pada Skema Jarak Klasifikasi 4 Kelas

No	Grup A	Grup B	Rerata Peringkat Grup A	Rerata Peringkat Grup B	Selisih	Nilai-P
1	Euclidean	Manhattan	1,54	1,46	0,08	0,3344

#### 4.11 Hasil Uji WSR pada Skema Uji Lanjut Tukey *HSD* dan WMW

Uji ini dilakukan untuk mengetahui apakah ada perbedaan signifikan antara penggunaan satu skema dengan skema lainnya untuk dua kelompok skema. Uji Wilcoxon Peringkat Bertanda pada skema uji lanjut berarti pengujian adakah perbedaan yang signifikan pada hasil klasifikasi dengan menggunakan Skema Tukey *HSD* dan Skema Wilcoxon Mann Whitney. Nilai RH untuk Skema Tukey *HSD* merupakan rerata untuk Skema F4 s.d. F6, sedangkan rerata Skema F7 s.d. F9 menjadi nilai RH untuk Wilcoxon Mann Whitney. Skema Hasil pengujian skema dengan hasil terbaik direkomendasikan untuk digunakan dalam penelitian selanjutnya.

##### 4.11.1 Uji WSR pada Skema Uji Lanjut Klasifikasi 2 Kelas

Hasil rerata peringkat menunjukkan bahwa Skema Wilcoxon Mann Whitney memiliki rerata peringkat tertinggi sebesar 1,70 dibanding skema Tukey *HSD* dengan rerata peringkat 1,30. Rerata peringkat dan varian peringkat dari masing-masing skema disajikan pada Tabel 4.32, sedangkan rincian proses pengujian pengaruh Skema Uji Lanjut terhadap blok uji diuraikan pada Lampiran 9.

Tabel 4.32 Rerata dan Varian Peringkat Skema Uji Lanjut (2 Kelas)

No	Metode Uji Lanjut	Rerata Peringkat	Varian Peringkat
1	Wilcoxon Mann Whitney	1,70	0,47
2	Tukey <i>HSD</i>	1,30	0,47

Hasil Uji Wilcoxon Peringkat Bertanda atas skema uji lanjut pada tingkat taraf nyata ( $\alpha=5\%$ ) menunjukkan bahwa tidak ada perbedaan signifikan antara penggunaan Skema Wilcoxon Mann Whitney dan Tukey *HSD* (Nilai-P =19,12%). Berdasarkan hasil Uji Wilcoxon Peringkat Bertanda dan rerata peringkat masing-masing skema, Skema Wilcoxon Mann Whitney direkomendasikan karena memiliki akurasi yang lebih tinggi dari Skema Manhattan.

Rincian hasil Uji Wilcoxon Peringkat Bertanda untuk skema uji lanjut pada klasifikasi 2 kelas disajikan pada Tabel 4.33.



Tabel 4.33 Hasil Uji WSR pada Skema Uji Lanjut Klasifikasi 2 Kelas

No	Grup A	Grup B	Rerata Peringkat Grup A	Rerata Peringkat Grup B	Selisih	Nilai-P
1	Tukey <i>HSD</i>	Wilcoxon Mann Whitney	1,30	1,70	0,40	0,1912

#### 4.11.2 Uji WSR pada Skema Uji Lanjut Klasifikasi 4 Kelas

Hasil rerata peringkat menunjukkan bahwa Skema Wilcoxon Mann Whitney memiliki rerata peringkat tertinggi sebesar 1,95 dibanding skema Tukey *HSD* dengan rerata peringkat 1,05. Rerata peringkat dan varian peringkat dari masing-masing skema disajikan pada Tabel 4.34, sedangkan rincian proses pengujian pengaruh Skema Uji Lanjut terhadap blok uji diuraikan pada Lampiran 10.

Tabel 4.34 Rerata dan Varian Peringkat Skema Uji Lanjut (4 Kelas)

No	Metode Uji Lanjut	Rerata Peringkat	Varian Peringkat
1	Wilcoxon Mann Whitney	1,95	0,22
2	Tukey <i>HSD</i>	1,05	0,22

Hasil Uji Wilcoxon Peringkat Bertanda atas skema uji lanjut pada tingkat taraf nyata ( $\alpha=5\%$ ) menunjukkan bahwa ada perbedaan signifikan antara penggunaan Skema Wilcoxon Mann Whitney dan Tukey *HSD* dengan Nilai-P sebesar 0,01% untuk keunggulan Skema Wilcoxon Mann Whitney. Berdasarkan hasil Uji Wilcoxon Peringkat Bertanda tersebut direkomendasikan penggunaan Skema Wilcoxon Mann Whitney untuk klasifikasi dengan 4 kelas label.

Rincian hasil Uji Wilcoxon Peringkat Bertanda untuk skema uji lanjut pada klasifikasi 4 kelas disajikan pada Tabel 4.35 (tanda \* pada Kolom Nilai-P menunjukkan ada perbedaan signifikan antar grup yang dibandingkan).

Tabel 4.35 Hasil Uji WSR pada Skema Uji Lanjut Klasifikasi 4 Kelas

No	Grup A	Grup B	Rerata Peringkat Grup A	Rerata Peringkat Grup B	Selisih	Nilai-P
1	Tukey <i>HSD</i>	Wilcoxon Mann Whitney	1,05	1,95	0,90	0,0001*

#### 4.12 Interpretasi Kinerja Klasifikasi Menggunakan AUC

Interpretasi AUC digunakan untuk pengambilan simpulan akhir apakah model klasifikasi data laporan keuangan untuk deteksi opini laporan keuangan sektor publik menggunakan berbagai skema fitur, termasuk fitur modifikasian dari sektor swasta layak diaplikasikan atau tidak. AUC tertinggi pada klasifikasi 2 kelas adalah sebesar 0,6630 dan AUC tertinggi untuk klasifikasi 4 kelas adalah sebesar 0,6156, kesemua model terbaik tersebut masuk ke kategori klasifikasi dengan kinerja yang buruk sesuai pendapat [44], sehingga bisa disimpulkan bahwa kesemua skema fitur memiliki kinerja yang buruk termasuk fitur modifikasian dari sektor swasta.

Pengujian lanjutan untuk klasifikasi 4 kelas dilakukan untuk membuktikan apakah skema awal yaitu jumlah data latih yang berbeda: WTP (30 data), WDP (10 data), TW (10 data), dan TMP (10 data) yang digunakan secara simultan lebih baik dari skema:

- Jumlah data latih yang berbeda, namun digunakan secara berurutan mulai dari WTP, WDP, TW dan terakhir TMP.
- Jumlah data latih yang sama ( $WTP=WDP=TW=TMP=10$  data) yang digunakan secara simultan.
- Jumlah data latih yang sama ( $WTP=WDP=TW=TMP=10$  data), namun digunakan secara berurutan mulai dari WTP, WDP, TW dan terakhir TMP.

Hasil pengujian lanjutan menunjukkan bahwa skema awal (skema nomor 1) memiliki AUC tertinggi dibanding skema yang lain. Hasil pengujian lanjutan tersebut diringkas pada Tabel 4.36 sebagai berikut:

Tabel 4.36 AUC dari Skema Pengujian Lanjutan untuk 4 Kelas

Skema	Jumlah data latih				Penggunaan Data Latih	AUC
	WTP	WDP	TW	TMP		
1	30	10	10	10	Semua secara simultan	0,6156
2	30	10	10	10	WTP saja	0,5000
3	30	10	10	10	WDP saja	0,5000
4	30	10	10	10	TW saja	0,5000
5	30	10	10	10	TMP saja	0,5000
6	10	10	10	10	Semua secara simultan	0,5311
7	10	10	10	10	WTP saja	0,5000
8	10	10	10	10	WDP saja	0,5000
9	10	10	10	10	TW saja	0,5000
10	10	10	10	10	TMP saja	0,5000

Rincian kualitas model klasifikasi dengan AUC tertinggi dan tingkat akurasi tertinggi menggunakan berbagai kombinasi skema normalisasi data (SND), fitur, nilai-K, dan jarak adalah sebagai berikut:

Tabel 4.37 Model Dengan AUC Tertinggi pada Klasifikasi 2 Kelas dan 4 Kelas

Klasifikasi	SND	Skema Fitur	Skema Nilai-K	Skema Jarak	Akurasi	AUC	Interpretasi AUC
2 Kelas	N4	F4 atau F7	B-NN	Euclidean	0,6805	0,6630	Buruk/ <i>poor</i>
4 Kelas	N1	F2	B-NN	Euclidean	0,5081	0,6156	Buruk/ <i>poor</i>

*Confusion matrix* dan perhitungan AUC dari skema terbaik untuk dua kelas disajikan pada Tabel 4.38, sedangkan untuk empat kelas disajikan pada Tabel 4.39 sebagai berikut:

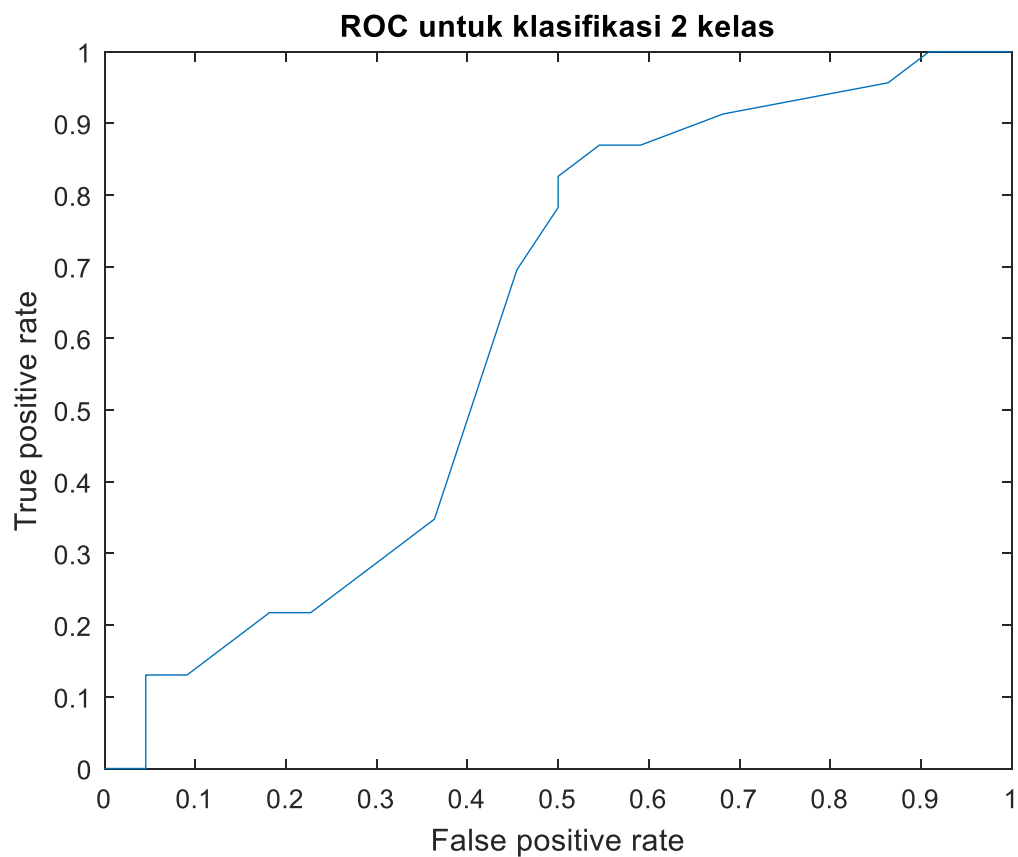
Tabel 4.38 *Confusion Matrix* dan AUC pada Klasifikasi 2 Kelas

Kondisi	Prediksi		TPR	FPR	AUC
	Baik	Buruk			
Baik	19	4	0,8261	0,5000	0,6630
Buruk	11	11	0,5000	0,1739	0,6630
Rerata Tertimbang			0,6667	0,3406	0,6630

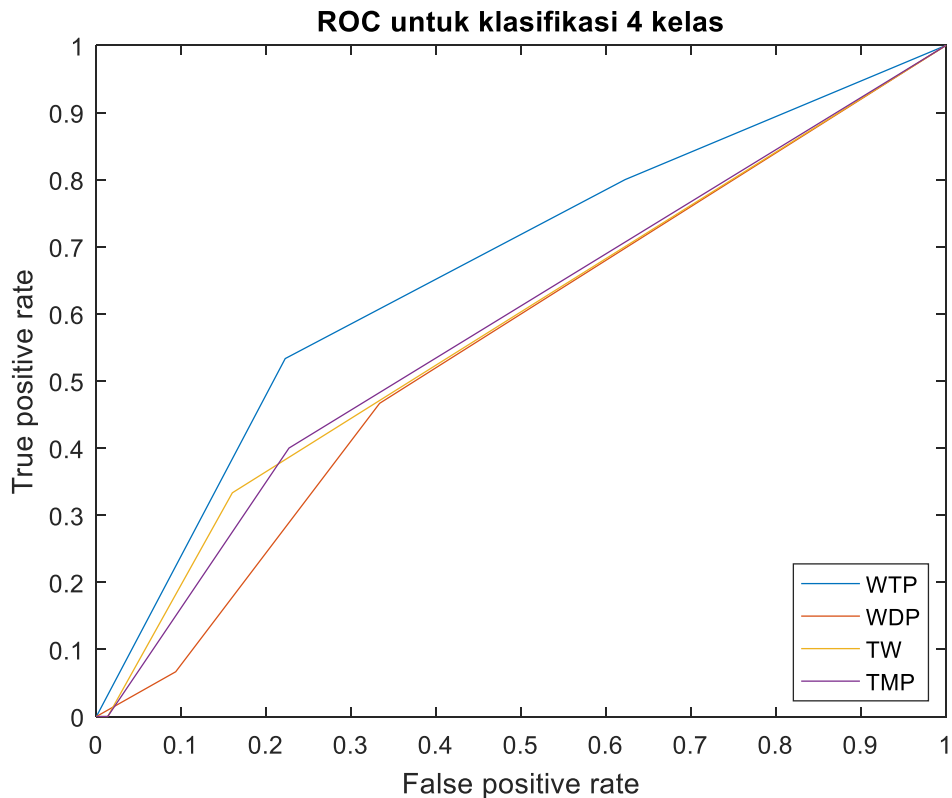
Tabel 4.39 *Confusion Matrix* dan AUC pada Klasifikasi 4 Kelas

	Prediksi				TPR	FPR	AUC
Kondisi	WTP	WDP	TW	TMP			
WTP	30	10	2	3	0,6667	0,3556	0,6556
WDP	5	6	1	3	0,4000	0,2267	0,5867
TW	6	4	3	2	0,2000	0,0800	0,5600
TMP	5	3	3	4	0,2667	0,1067	0,5800
Rerata Tertimbang					0,4778	0,2467	0,6156

Kurva ROC untuk skema terbaik klasifikasi 2 kelas disajikan pada Gambar 4.1 sedangkan untuk empat kelas disajikan pada Gambar 4.2 sebagai berikut:



Gambar 4.1 Kurva ROC untuk Klasifikasi 2 Kelas



Gambar 4.2 Kurva ROC untuk Klasifikasi 4 Kelas

#### 4.13 Skema Sistem Alarm (*Early Warning System*) Anomali Opini

Penelitian ini diharapkan dapat memberikan alarm jika terjadi anomali opini atas Laporan Hasil Pemeriksaan atas Laporan Keuangan (LHP LK) suatu entitas. Skema sistem diperoleh dari hasil uji statistik atas kombinasi skema (SND, fitur, Nilai-K, Jarak, dan uji lanjut untuk pemilihan fitur) yang merupakan rerata dari seluruh percobaan yang dilakukan.

##### 4.13.1 Skema Sistem Alarm Klasifikasi 2 Kelas

Hasil uji statistik untuk klasifikasi 2 kelas menunjukkan bahwa:

- Tidak ada perbedaan signifikan antara kinerja Skema Normalisasi Data (SND) nonlinier dan linier. Akan tetapi, Skema N3 (normalisasi menggunakan Nilai- $Z$ /  $Z$ -score dari data) memiliki rerata peringkat tertinggi dibanding skema lainnya.

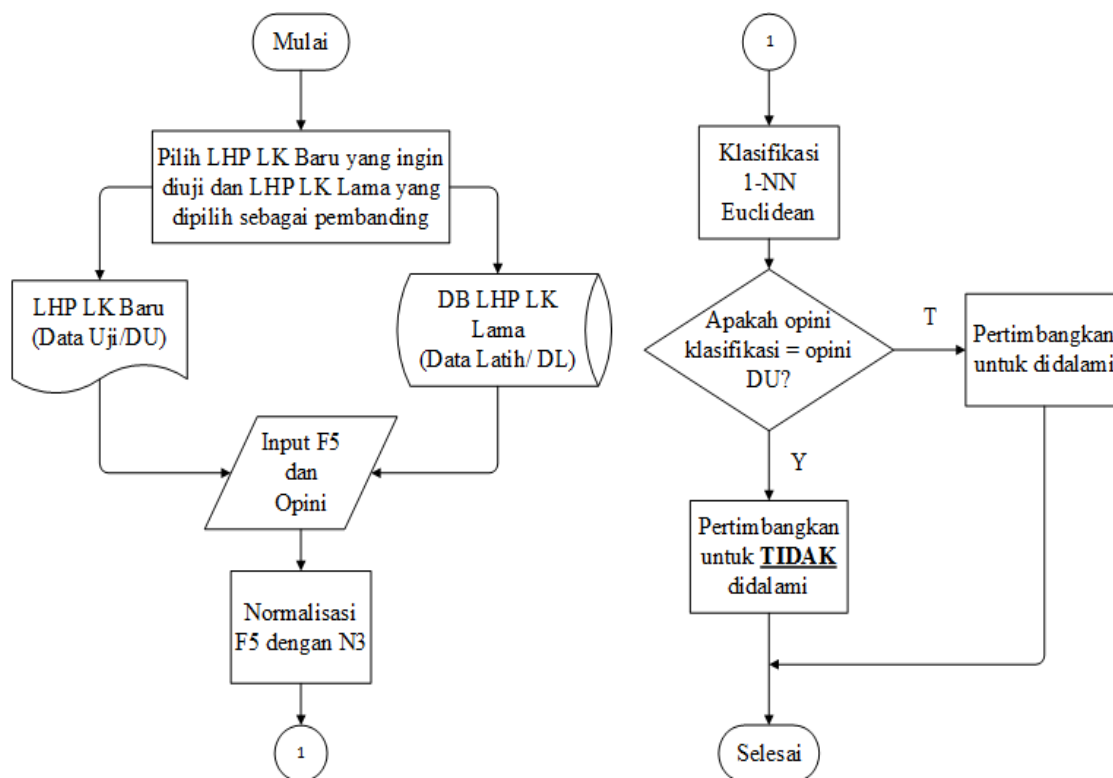
- b. Skema F5 atau F8 (fitur signifikan dengan Nilai Log10 terkecil sesuai Uji Wilcoxon Mann Whitney) unggul secara signifikan dibandingkan dengan Skema F1 (fitur modifikasian dari sektor swasta).
- c. Skema 1-NN memiliki rerata peringkat tertinggi dibanding Skema B-NN.
- d. Jarak Euclidean dan Manhattan memiliki kinerja yang sama pada klasifikasi 2 kelas.
- e. Tidak ada perbedaan signifikan antara kinerja uji lanjut Wilcoxon Mann Whitney (WMW) dan Tukey *HSD*, namun WMW memiliki rerata peringkat lebih tinggi dibanding Tukey *HSD*.

Berdasarkan hasil uji statistik tersebut, maka disusunlah skema sistem alarm untuk dua kelas, adapun tabel bantu untuk memahami kode pada gambar sistem alarm untuk anomali opini LHP LK untuk dua kelas (WTP dan selain WTP) disajikan Tabel 4.40 berikut:

Tabel 4.40 Tabel Bantu Skema Sistem Alarm Klasifikasi 2 Kelas

No	Kode	Uraian
1	N3	Normalisasi menggunakan Skor-Z/ <i>Z-Score</i> sesuai Persamaan (3.3)
2	F5 atau F8	Fitur-fitur pilihan sebagai berikut: Nomor 13= PAD/Total Aset Nomor 15= PAD/Aset Tetap Nomor 36= PAD/Pendapatan Transfer
3	1-NN	Klasifikasi menggunakan Nilai-K =1
4	LHP LK	Laporan Hasil Pemeriksaan atas Laporan Keuangan

Gambar sistem alarm anomali opini LHP LK untuk klasifikasi dua kelas (WTP dan selain WTP) disajikan pada Gambar 4.3 berikut:



Gambar 4.3 Skema Sistem Alarm Klasifikasi 2 Kelas

#### 4.13.2 Skema Sistem Alarm Klasifikasi 4 Kelas

Hasil uji statistik untuk klasifikasi 4 kelas menunjukkan bahwa:

- Tidak ada perbedaan signifikan antara kinerja Skema Normalisasi Data (SND) nonlinier dan linier. Akan tetapi, Skema N4 (normalisasi menggunakan Eksponensial Nilai-Z/ *Z-score* dari data) memiliki rerata peringkat tertinggi dibanding skema lainnya.
- Skema F2 (fitur orisinal dari sektor publik terkait dengan belanja pemerintah) unggul secara signifikan dibandingkan dengan Skema F1 (fitur modifikasi dari sektor swasta).
- Skema B-NN unggul secara signifikan dibanding Skema 1-NN.
- Tidak ada perbedaan signifikan antara kinerja Jarak Euclidean dan Manhattan, namun Euclidean memiliki rerata peringkat lebih tinggi.

- e. Uji lanjut Wilcoxon Mann Whitney (WMW) unggul secara signifikan dibandingkan dengan Tukey *HSD*.

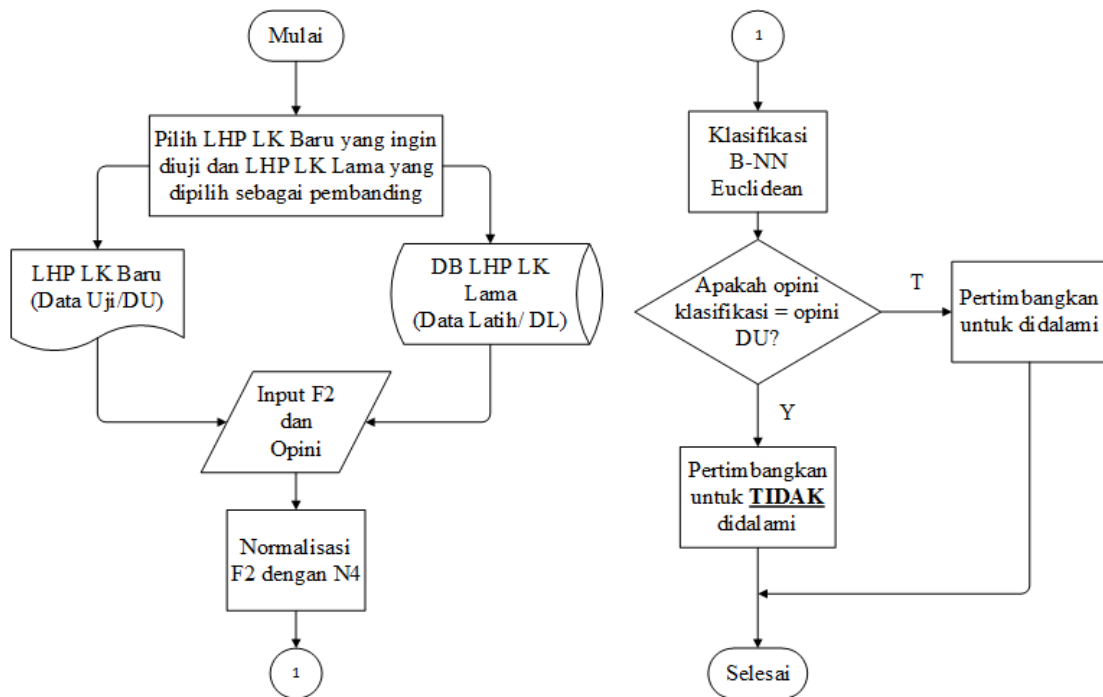
Berdasarkan hasil uji statistik tersebut, maka disusunlah skema sistem alarm untuk empat kelas, adapun tabel bantu untuk memahami kode pada gambar sistem alarm untuk anomali opini LHP LK untuk dua kelas (WTP dan selain WTP) disajikan Tabel 4.41 berikut:

Tabel 4.41 Tabel Bantu Skema Sistem Alarm Klasifikasi 4 Kelas

<b>No</b>	<b>Kode</b>	<b>Uraian</b>
1	N4	Normalisasi menggunakan eksponensial Skor-Z/ <i>Z-Score</i> sesuai Persamaan (3.4)
2	F2	Fitur asli dari sektor publik sebagai berikut: Nomor 30= Belanja Modal/Perubahan Nilai Aset Tetap Nomor 31= Belanja Barang Jasa/Perubahan Nilai Persediaan Nomor 32= Belanja Pegawai/Total Belanja Nomor 33= Belanja Modal/Total Belanja Nomor 34= Belanja Hibah/Total Belanja Nomor 35= Belanja Bantuan Sosial/Total Belanja Nomor 36= PAD/Pendapatan Transfer Nomor 37= Zona Wilayah (Barat atau Timur) Nomor 38= Jenis Wilayah Administratif (Provinsi, Kabupaten, atau Kota)
3	B-NN	Klasifikasi menggunakan Nilai-K terbaik menggunakan validasi 10 lipatan pada data latih dengan akurasi tertinggi
4	LHP LK	Laporan Hasil Pemeriksaan atas Laporan Keuangan



Gambar sistem alarm anomali opini LHP LK untuk klasifikasi 4 kelas (WTP, WDP, TW dan TMP) disajikan pada Gambar 4.4 berikut:



Gambar 4.4 Skema Sistem Alarm Klasifikasi 4 Kelas

*Halaman ini sengaja dikosongkan*

## **BAB 5**

### **PENUTUP**

#### **5.1 Kesimpulan**

Berdasarkan hasil penelitian yang telah diperoleh dapat disimpulkan bahwa:

1. Metode klasifikasi yang dihasilkan penelitian ini, baik untuk 2 kelas maupun 4 kelas, tergolong klasifikasi yang buruk karena memiliki nilai AUC tertinggi pada rentang 0,60 s.d. 0,70. Perlu penelitian lebih lanjut sebelum dapat diaplikasikan pada kondisi nyata.
2. Penelitian ini menghasilkan AUC tertinggi untuk 2 kelas sebesar 66,30% dengan penggunaan kombinasi normalisasi data menggunakan eksponensial *Z-Score*, jarak euclidean, fitur nomor 2, 4, 6, 7, 11, 13, 15, 24, 26, 35, 36 dan nilai-k terbaik/k=2. Jika menggunakan uji statistik, maka skema terbaik untuk 2 kelas adalah kombinasi normalisasi menggunakan *Z-Score*, jarak euclidean, fitur nomor 13, 15, 36, dan nilai-k=1.
3. Penelitian ini menghasilkan AUC tertinggi untuk 4 kelas sebesar 61,56% dengan penggunaan kombinasi normalisasi data dengan skop (0 s.d. 1), jarak euclidean, fitur nomor 30 s.d. 38 dan nilai-k terbaik/k=2. Jika menggunakan uji statistik, maka skema terbaik untuk 4 kelas adalah kombinasi normalisasi menggunakan eksponensial *Z-Score*, jarak euclidean, fitur nomor 30 s.d. 38 dan nilai-k terbaik/k=2.

#### **5.2 Saran**

Penelitian ini menggunakan data dalam jumlah kecil (150 data) serta menggunakan 29 fitur termodifikasi dari sektor swasta ke sektor publik. Faktor-faktor tersebut ditengarai menjadi penyebab rendahnya kinerja model klasifikasi yang dihasilkan. Selain itu, penggunaan fitur termodifikasi dari sektor swasta yang cenderung “*profit-oriented*” juga berisiko pada biasanya hasil penelitian pada sektor publik yang cenderung “*serviced-oriented/cost-oriented*”. Sembilan fitur

termodifikasi dengan karakteristik sektor swasta yang kurang tepat diaplikasikan untuk sektor publik disajikan pada Tabel 5.1 berikut:

Tabel 5.1 Fitur Publik Termodifikasi Berkarakter Sektor Swasta

No	Fitur Swasta	Fitur Termodifikasi
1	<i>Net profit</i>	SILPA (Sisa Lebih Pembiayaan Anggaran)
2	<i>Primary business income</i>	PAD (pendapatan asli daerah)
3	<i>Inventory/Primary business income</i>	Kas Tunai dan Kas Bank/PAD
4	<i>Net profit/Total assets</i>	SILPA/Total Aset
5	<i>Net profit/Primary business income</i>	SILPA/PAD
6	<i>Accounts receivable/Primary business income</i>	Piutang/PAD
7	<i>Primary business income/Total assets</i>	PAD/Total Aset
8	<i>Primary business income/Fixed assets</i>	PAD/Aset Tetap
9	<i>Primary business income/Last year's primary business income</i>	PAD/PAD Tahun Lalu

Berdasarkan uraian tersebut, penelitian selanjutnya dapat melakukan perbaikan penelitian dengan cara:

1. Peningkatan jumlah data laporan keuangan yang digunakan dalam penelitian.
2. Eksplorasi dan penggunaan lebih banyak fitur asli dari sektor publik, baik data kuantitatif maupun kualitatif, baik data keuangan maupun nonkeuangan.
3. Pengembangan model penelitian lebih lanjut untuk deteksi penyebab kualifikasi atau deteksi temuan audit (*audit findings*) yang berdampak pada akun laporan keuangan.
4. Penggunaan dan perbandingan kinerja antar lebih dari satu metode kecerdasan buatan untuk mencari hasil klasifikasi terbaik.
5. Penggunaan studi kasus untuk memperkuat hasil penelitian.

## DAFTAR PUSTAKA

- [1] K. M. Dye, "Corruption and Fraud Detection by Supreme Audit Institutions," in *Performance Accountability and Combating Corruption Audit Institutions*, Washington, DC, The World Bank, 2007, pp. 303-321.
- [2] Association of Certified Fraud Examiners (ACFE), "Report to the Nation on Occupational Fraud and Abuse 2016 Global Fraud Study," ACFE, Austin, 2016.
- [3] Ikatan Akuntan Indonesia, Kode Etik Akuntan Profesional, Jakarta: Ikatan Akuntan Indonesia, 2016.
- [4] International Ethics Standards Board for Accountants, Handbook of the Code of Ethics for Professional Accountants, New York: International Federation of Accountants, 2015.
- [5] C. Tikasari, "Pengaruh Penerapan Nilai Dasar Kode Etik BPK-RI Terhadap Kinerja Auditor Pemerintah (Studi Pada Auditor BPK-RI Perwakilan Provinsi Jawa Barat)," Universitas Pendidikan Indonesia, Bandung, 2013.
- [6] D. A. Trihapsari and I. Anisykurlillah, "Pengaruh Etika, Independensi, Pengalaman Audit dan Premature Sign Off Terhadap Kualitas Audit," *Accounting Analysis Journal*, vol. 5, no. 1, pp. 1-7, 2016.
- [7] Amrizal, "Analisis Kritis Pelanggaran Kode Etik Profesi Akuntan Publik di Indonesia," *Jurnal Liquidity*, vol. 3, no. 1, pp. 36-43, 2014.
- [8] D. Satava, C. Caldwell and L. Richards, "Ethics and the Auditing Culture: Rethinking the Foundation of Accounting and Auditing," *Journal of Business Ethics*, vol. 64, p. 271–284, 2006.
- [9] Badan Pemeriksa Keuangan Republik Indonesia, Ikhtisar Hasil Pemeriksaan Semester I Tahun 2015, Jakarta, 2015.
- [10] Badan Pemeriksa Keuangan Republik Indonesia, Ikhtisar Hasil Pemeriksaan Semester I Tahun 2014, Jakarta, 2014.
- [11] Badan Pemeriksa Keuangan Republik Indonesia, Ikhtisar Hasil Pemeriksaan Semester II Tahun 2014, Jakarta, 2014.
- [12] Badan Pemeriksa Keuangan Republik Indonesia, Ikhtisar Hasil Pemeriksaan Semester II Tahun 2015, Jakarta, 2015.
- [13] Badan Pemeriksa Keuangan Republik Indonesia, Ikhtisar Hasil Pemeriksaan Semester I Tahun 2016, Jakarta, 2016.
- [14] A. A. Baldwin, C. E. Brown and B. S. Trinkle, "Opportunities For Artificial Intelligence Development In The Accounting Domain: The Case For

Auditing,” *Intelligent Systems In Accounting, Finance And Management*, pp. Vol. 14 Hal.77-86, 2006.

- [15] ndr/nrl; , “detiknews: BPK Jamin Tindak Tegas Oknum yang 'Jual Beli' Opini Hasil Audit,” detik.com, 2010. [Online]. Available: <http://news.detik.com/berita/1384815/bpk-jamin-tindak-tegas-oknum-yang-jual-beli-opini-hasil-audit>. [Accessed 20 Desember 2016 Pukul 08.25 WIB].
- [16] Haryadi, Dedi ; Kompas (Rabu, 11 Mei 2016 Halaman 7), “BPK RI-DARI MEDIA-Korupsi Audit,” 2016. [Online]. Available: <http://www.bpk.go.id/news/korupsi-audit>. [Accessed 20 Desember 2016 Pukul 08.20 WIB].
- [17] C. Gaganis, F. Pasiouras and M. Doumpos, “Probabilistic neural networks for the identification of qualified audit opinions,” *Expert Systems with Applications*, pp. Vol. 32 Hal. 114-124. Elsevier, 2007.
- [18] C. Gaganis, F. Pasiouras, C. Spathis and C. Zopounidis, “A comparison of nearest neighbours, discriminant and logit models for auditing decisions,” *Intelligent Systems in Accounting, Finance & Management*, vol. 15, no. 1-2, pp. 23-40, 2007.
- [19] C. Gaganis, “Classification Techniques For The Identification Of Falsified Financial Statements: A Comparative Analysis,” *Intelligent Systems In Accounting, Finance And Management*, pp. Vol. 16, Hal. 207-229, 2009.
- [20] S. Chen, “Detection Of Fraudulent Financial Statements Using The Hybrid Data Mining Approach,” *Springer Plus*, p. Vol. 5, 2016.
- [21] E. Ngai, Y. Hu, Y. Wong, Y. Chen and X. Sun, “The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature,” *Decision Support Systems*, pp. Vol. 50 Hal. 559-569, 2011.
- [22] P. Ravisankar, V. Ravi, G. R. Rao and I. Bose, “Detection Of Financial Statement Fraud And Feature Selection Using Data Mining Techniques,” *Decision Support Systems*, pp. Vol. 50 Hal. 491-500, 2011.
- [23] S. Thiprungsri and M. A. Vasarhelyi, “Cluster Analysis for Anomaly Detection in Accounting Data: An Audit Approach,” *The International Journal of Digital Accounting Research*, vol. 11, pp. 69 - 84, 2011.
- [24] S. Yang, J. Huang, D. Zhou, H. Zha and C. L. Giles, “IKNN: Informative K-Nearest Neighbor Pattern Classification,” in *Proceedings of the 11th European Conference on Principles and Practice of Knowledge Discovery in Databases*, Warsaw, Poland, Springer-Verlag, 2007, pp. 248-264.
- [25] E. Prasetyo, *Data Mining Mengolah Data Menjadi Informasi Menggunakan Matlab*, Yogyakarta: Andi Publisher, 2014.

- [26] J. Tamouk and F. Allahakbari, "A comparison among accuracy of KNN, PNN, KNCN, DANN and NFL," *IJCSI International Journal of Computer Science Issues*, vol. 9, no. 3 No. 1, pp. 319-322, 2012.
- [27] A. Statnikov, C. F. Aliferis, I. Tsamardinos, D. Hardin and S. Levy, "A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis," *Bioinformatics*, vol. 21, pp. 631-643, 2005.
- [28] Pemerintah Republik Indonesia, Undang-Undang Republik Indonesia Nomor 17 Tahun 2003 Tentang Keuangan Negara, 2013.
- [29] Badan Pemeriksa Keuangan Republik Indonesia, Keputusan BPK RI Nomor 4/K/I-XIII.2/9/2012 Tentang Buletin Teknis Nomor 01 Pelaporan Hasil Pemeriksaan Atas Laporan Keuangan, BPK RI. Jakarta, 2012.
- [30] L. Debby, "debbylinia.blogspot.co.id," 27 Agustus 2014. [Online]. Available: <http://debbylinia.blogspot.co.id/2014/08/makalah-ekonomi-politik-pemerataan.html>. [Accessed 15 Maret 2017].
- [31] U. Fayyad, G. P. Shapiro and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," *Fall*, pp. Vol. 17 No. 3 Hal. 37-54, 1996.
- [32] E. Prasetyo, Data Mining : Konsep dan Aplikasi Menggunakan Matlab, Penerbit Andi, Yogyakarta, 2012.
- [33] S. Muliadinata, "sharewy.blogspot.co.id," 2013. [Online]. Available: [http://sharewy.blogspot.co.id/2013/04/algoritma-k-nearest-neighbor-knn\\_16.html](http://sharewy.blogspot.co.id/2013/04/algoritma-k-nearest-neighbor-knn_16.html). [Accessed 15 Maret 2017].
- [34] F. Wilcoxon, "Individual Comparisons by Ranking Methods," *Biometrics Bulletin*, vol. 1, no. 6, pp. 80-83, 1945.
- [35] M. S. Dahlan, Statistik untuk Kedokteran dan Kesehatan Deskriptif, Bivariat, dan Multivariat Dilengkapi Aplikasi dengan Menggunakan SPSS, Jakarta: Salemba Medika, 2001.
- [36] S. Arini, "Uji Wilcoxon-Mann Whitney Kasus 2 Sampel Independen," 31 Januari 2014. [Online]. Available: <http://arini2992.blogspot.com/2014/01/uji-wilcoxon-mann-whitney-kasus-2.html>. [Accessed 2 April 2017].
- [37] W. H. Kruskal and W. A. Wallis, "Use of Ranks in One-Criterion Variance Analysis," *Journal of the American Statistical Association*, vol. 47, no. 260, pp. 583-621, 1952.
- [38] The MathWorks, Inc., "MathWorks Documentation Multiple Comparisons," 2006, [Online]. Available: <https://www.mathworks.com/help/stats/multiple-comparisons.html#bum7ugv-1>. [Accessed 6 April 2017].

- [39] M. Friedman, "The Use of Ranks to Avoid The Assumption of Normality Implicit in The Analysis of Variance," *Journal of The American Statistical Association*, vol. 32, no. 200, pp. 675-701, 1937.
- [40] Minitab Inc., "Minitab Express Support - Methods and Formulas for Friedman Test," 2016. [Online]. Available: <http://support.minitab.com/en-us/minitab-express/1/help-and-how-to/modeling-statistics/anova/how-to/friedman-test/methods-and-formulas/methods-and-formulas/>. [Accessed 2 April 2017].
- [41] W. Solidayah, S. Sunendiari and L. Wachidah, "Uji Modifikasi Peringkat Bertanda Wilcoxon Untuk Masalah Dua Sampel Berpasangan," in *Prosiding Statistika Seminar Penelitian Sivitas Akademika Unisba (SPeSIA)*, Bandung, 2015.
- [42] A. Saifudin and R. S. Wahono, "Pendekatan Level Data untuk Menangani Ketidakseimbangan Kelas pada Prediksi Cacat Software," *Journal of Software Engineering*, vol. 1, no. 2, pp. 76-85, 2015.
- [43] N. Chinchor, "MUC-4 Evaluation Metric," 1992. [Online]. Available: <http://acl-arc.comp.nus.edu.sg/archives/acl-arc-090501d4/data/pdf/anthology-PDF/M/M92/M92-1002.pdf>. [Accessed 20 Maret 2017].
- [44] F. Gorunescu, *Data Mining Concepts, Models and Techniques*, Verlag Berlin Heidelberg: Springer, 2011.
- [45] I. Guyon and A. Elisseeff, "An introduction to variable and feature selection," *Journal of machine learning research*, vol. 3, pp. 1157-1182, 2003.
- [46] R. Kohavi, "A Study of Cross Validation and Bootstrap for Accuracy Estimation and Model Selection," in *Proceedings of the 14th International Joint Conference on Artificial Intelligence - Volume 2*, San Francisco, CA, USA, Morgan Kaufmann Publishers Inc., 1995, pp. 1137-1143.



## LAMPIRAN

Lampiran 1 Blok Pengujian Pengaruh Normalisasi pada Rerata Harmonik AUC dan *F-Measure/RH* (Klasifikasi 2 Kelas)

Blok	Blok Penelitian (Kombinasi Skema Lain)			Pengaruh Skema Normalisasi Data (RH)					Peringkat				
	SNN	Fitur	Jarak	N0	N1	N2	N3	N4	N0	N1	N2	N3	N4
1	1-NN	F1	euclidean	0,4887	0,5098	0,5098	0,5076	0,4593	2	4,5	4,5	3	1
2	1-NN	F1	manhattan	0,5333	0,5942	0,5942	0,5732	0,5311	2	4,5	4,5	3	1
3	1-NN	F2	euclidean	0,4629	0,5778	0,5778	0,5779	0,5779	1	2,5	2,5	4,5	4,5
4	1-NN	F2	manhattan	0,4419	0,6223	0,6223	0,6002	0,6002	1	4,5	4,5	2,5	2,5
5	1-NN	F3	euclidean	0,4887	0,6433	0,6433	0,5988	0,5543	1	4,5	4,5	3	2
6	1-NN	F3	manhattan	0,5333	0,5969	0,5969	0,5732	0,5942	1	4,5	4,5	2	3
7	1-NN	F4	euclidean	0,5720	0,5768	0,5768	0,5757	0,5311	2	4,5	4,5	3	1
8	1-NN	F4	manhattan	0,5510	0,6224	0,6224	0,5757	0,5757	1	4,5	4,5	2,5	2,5
9	1-NN	F5	euclidean	0,5984	0,5779	0,5779	0,6223	0,6443	3	1,5	1,5	4	5
10	1-NN	F5	manhattan	0,6223	0,5557	0,5557	0,5997	0,5997	5	1,5	1,5	3,5	3,5
11	1-NN	F6	euclidean	0,5536	0,5768	0,5768	0,5321	0,5536	2,5	4,5	4,5	1	2,5
12	1-NN	F6	manhattan	0,5778	0,5551	0,5551	0,5321	0,5321	5	3,5	3,5	1,5	1,5
13	1-NN	F7	euclidean	0,5720	0,5768	0,5768	0,5757	0,5311	2	4,5	4,5	3	1
14	1-NN	F7	manhattan	0,5510	0,6224	0,6224	0,5757	0,5757	1	4,5	4,5	2,5	2,5
15	1-NN	F8	euclidean	0,5984	0,5779	0,5779	0,6223	0,6443	3	1,5	1,5	4	5
16	1-NN	F8	manhattan	0,6223	0,5557	0,5557	0,5997	0,5997	5	1,5	1,5	3,5	3,5
17	1-NN	F9	euclidean	0,5554	0,4668	0,4668	0,5112	0,5112	5	1,5	1,5	3,5	3,5
18	1-NN	F9	manhattan	0,6217	0,5333	0,5333	0,5335	0,5335	5	1,5	1,5	3,5	3,5
19	B-NN	F1	euclidean	0,4659	0,5557	0,5999	0,5732	0,6203	1	2	4	3	5
20	B-NN	F1	manhattan	0,4427	0,5779	0,6217	0,5311	0,5772	1	4	5	2	3
21	B-NN	F2	euclidean	0,4629	0,5778	0,5778	0,5779	0,5779	1	2,5	2,5	4,5	4,5

Lampiran 1 Blok Pengujian Pengaruh Normalisasi pada Rerata Harmonik AUC dan *F-Measure*/RH (Klasifikasi 2 Kelas) - Lanjutan

Blok	Blok Penelitian (Kombinasi Skema Lain)			Pengaruh Skema Normalisasi Data (RH)					Peringkat				
	SNN	Fitur	Jarak	N0	N1	N2	N3	N4	N0	N1	N2	N3	N4
22	B-NN	F2	manhattan	0,4419	0,6223	0,6223	0,6002	0,6002	1	4,5	4,5	2,5	2,5
23	B-NN	F3	euclidean	0,4659	0,5779	0,5779	0,5988	0,5988	1	2,5	2,5	4,5	4,5
24	B-NN	F3	manhattan	0,5333	0,5969	0,5969	0,5732	0,5942	1	4,5	4,5	2	3
25	B-NN	F4	euclidean	0,4659	0,5778	0,6002	0,5543	0,6601	1	3	4	2	5
26	B-NN	F4	manhattan	0,4874	0,5768	0,5768	0,6315	0,6150	1	2,5	2,5	5	4
27	B-NN	F5	euclidean	0,6173	0,5749	0,5749	0,6223	0,6173	3,5	1,5	1,5	5	3,5
28	B-NN	F5	manhattan	0,5961	0,5749	0,5749	0,6173	0,6173	3	1,5	1,5	4,5	4,5
29	B-NN	F6	euclidean	0,4608	0,4659	0,4850	0,4850	0,4850	1	2	4	4	4
30	B-NN	F6	manhattan	0,4608	0,4608	0,4608	0,4850	0,4850	2	2	2	4,5	4,5
31	B-NN	F7	euclidean	0,4659	0,5778	0,6002	0,5543	0,6601	1	3	4	2	5
32	B-NN	F7	manhattan	0,4874	0,5768	0,5768	0,6315	0,6150	1	2,5	2,5	5	4
33	B-NN	F8	euclidean	0,6173	0,5749	0,5749	0,6223	0,6173	3,5	1,5	1,5	5	3,5
34	B-NN	F8	manhattan	0,5961	0,5749	0,5749	0,6173	0,6173	3	1,5	1,5	4,5	4,5
35	B-NN	F9	euclidean	0,5059	0,5359	0,5929	0,5885	0,5359	1	2,5	5	4	2,5
36	B-NN	F9	manhattan	0,5088	0,4890	0,5424	0,5018	0,5018	4	1	5	2,5	2,5
Rerata									2,18	2,90	3,28	3,33	3,31
Simp. Baku									1,46	1,30	1,37	1,11	1,26

Lampiran 2 Blok Pengujian Pengaruh Normalisasi pada Rerata Harmonik AUC dan *F-Measure* (Klasifikasi 4 Kelas)

Blok	Blok Penelitian (Kombinasi Skema Lain)			Pengaruh Skema Normalisasi Data (RH)					Peringkat				
	SNN	Fitur	Jarak	N0	N1	N2	N3	N4	N0	N1	N2	N3	N4
1	1-NN	F1	euclidean	0,4260	0,4129	0,4129	0,4591	0,4492	3	1,5	1,5	5	4
2	1-NN	F1	manhattan	0,4502	0,4096	0,4096	0,4242	0,4293	5	1,5	1,5	3	4
3	1-NN	F2	euclidean	0,3284	0,5352	0,5352	0,5129	0,5296	1	4,5	4,5	2	3
4	1-NN	F2	manhattan	0,3519	0,5212	0,5212	0,4899	0,5341	1	3,5	3,5	2	5
5	1-NN	F3	euclidean	0,4260	0,4685	0,4685	0,4722	0,4653	1	3,5	3,5	5	2
6	1-NN	F3	manhattan	0,4502	0,4646	0,4646	0,4660	0,4738	1	2,5	2,5	4	5
7	1-NN	F4	euclidean	0,3902	0,4147	0,4147	0,4214	0,4466	1	2,5	2,5	4	5
8	1-NN	F4	manhattan	0,4246	0,4505	0,4505	0,4781	0,4865	1	2,5	2,5	4	5
9	1-NN	F5	euclidean	0,4001	0,4001	0,4001	0,4001	0,4066	2,5	2,5	2,5	2,5	5
10	1-NN	F5	manhattan	0,4001	0,4001	0,4001	0,4001	0,4066	2,5	2,5	2,5	2,5	5
11	1-NN	F6	euclidean	0,3696	0,3584	0,3584	0,3565	0,3613	5	2,5	2,5	1	4
12	1-NN	F6	manhattan	0,3386	0,3247	0,3247	0,3425	0,3396	3	1,5	1,5	5	4
13	1-NN	F7	euclidean	0,4067	0,4314	0,4314	0,4344	0,4665	1	2,5	2,5	4	5
14	1-NN	F7	manhattan	0,4772	0,4949	0,4949	0,4761	0,4995	2	3,5	3,5	1	5
15	1-NN	F8	euclidean	0,4204	0,4204	0,4204	0,4204	0,4459	2,5	2,5	2,5	2,5	5
16	1-NN	F8	manhattan	0,4257	0,4282	0,4282	0,4180	0,4252	3	4,5	4,5	1	2
17	1-NN	F9	euclidean	0,3902	0,4147	0,4147	0,4214	0,4466	1	2,5	2,5	4	5
18	1-NN	F9	manhattan	0,4246	0,4505	0,4505	0,4781	0,4865	1	2,5	2,5	4	5
19	B-NN	F1	euclidean	0,4260	0,4186	0,4186	0,4213	0,4026	5	2,5	2,5	4	1
20	B-NN	F1	manhattan	0,4502	0,4375	0,4375	0,4491	0,4551	4	1,5	1,5	3	5
21	B-NN	F2	euclidean	0,4000	0,5352	0,5352	0,5129	0,5296	1	4,5	4,5	2	3
22	B-NN	F2	manhattan	0,4000	0,5378	0,5212	0,5346	0,5309	1	5	2	4	3

Lampiran 2 Blok Pengujian Pengaruh Normalisasi pada Rerata Harmonik AUC dan *F-Measure* (Klasifikasi 4 Kelas) - Lanjutan

Blok	Blok Penelitian (Kombinasi Skema Lain)			Pengaruh Skema Normalisasi Data (RH)					Peringkat				
	SNN	Fitur	Jarak	N0	N1	N2	N3	N4	N0	N1	N2	N3	N4
23	B-NN	F3	euclidean	0,4260	0,4201	0,4201	0,4293	0,4264	3	1,5	1,5	5	4
24	B-NN	F3	manhattan	0,4502	0,4000	0,4000	0,4643	0,4086	4	1,5	1,5	5	3
25	B-NN	F4	euclidean	0,4256	0,5001	0,5001	0,4768	0,4694	1	4,5	4,5	3	2
26	B-NN	F4	manhattan	0,4246	0,4587	0,4464	0,4757	0,4563	1	4	2	5	3
27	B-NN	F5	euclidean	0,4140	0,4140	0,4140	0,4140	0,4140	3	3	3	3	3
28	B-NN	F5	manhattan	0,4140	0,4140	0,4140	0,4140	0,4140	3	3	3	3	3
29	B-NN	F6	euclidean	0,4172	0,4083	0,4112	0,4106	0,4106	5	1	4	2,5	2,5
30	B-NN	F6	manhattan	0,4142	0,4017	0,3819	0,3889	0,3889	5	4	1	2,5	2,5
31	B-NN	F7	euclidean	0,4067	0,5001	0,5001	0,4757	0,4591	1	4,5	4,5	3	2
32	B-NN	F7	manhattan	0,4607	0,4487	0,4487	0,4606	0,4444	5	2,5	2,5	4	1
33	B-NN	F8	euclidean	0,4204	0,4832	0,4832	0,4204	0,4079	2,5	4,5	4,5	2,5	1
34	B-NN	F8	manhattan	0,4120	0,4734	0,4079	0,4734	0,4734	2	4	1	4	4
35	B-NN	F9	euclidean	0,4256	0,5001	0,4842	0,4768	0,4694	1	5	4	3	2
36	B-NN	F9	manhattan	0,4246	0,4756	0,4587	0,4641	0,4563	1	5	3	4	2
Rerata									2,39	3,07	2,76	3,31	3,47
Simp. Baku									1,51	1,19	1,08	1,17	1,36

Lampiran 3 Blok Pengujian Pengaruh Fitur pada Rerata Harmonik AUC dan *F-Measure* (Klasifikasi 2 Kelas)

Blok	Blok Penelitian			Pengaruh Skema Fitur									Peringkat								
	SNN	SND	Jarak	F1	F2	F3	F4	F5	F6	F7	F8	F9	F1	F2	F3	F4	F5	F6	F7	F8	F9
1	1-NN	N0	euclidean	0,49	0,46	0,49	0,57	0,60	0,55	0,57	0,60	0,56	2,5	1	2,5	6,5	8,5	4	6,5	8,5	5
2	1-NN	N0	manhattan	0,53	0,44	0,53	0,55	0,62	0,58	0,55	0,62	0,62	2,5	1	2,5	4,5	8,5	6	4,5	8,5	7
3	1-NN	N1	euclidean	0,51	0,58	0,64	0,58	0,58	0,58	0,58	0,58	0,47	2	6	9	4	7,5	4	4	7,5	1
4	1-NN	N1	manhattan	0,59	0,62	0,60	0,62	0,56	0,56	0,62	0,56	0,53	5	7	6	8,5	3,5	2	8,5	3,5	1
5	1-NN	N2	euclidean	0,51	0,58	0,64	0,58	0,58	0,58	0,58	0,58	0,47	2	6	9	4	7,5	4	4	7,5	1
6	1-NN	N2	manhattan	0,59	0,62	0,60	0,62	0,56	0,56	0,62	0,56	0,53	5	7	6	8,5	3,5	2	8,5	3,5	1
7	1-NN	N3	euclidean	0,51	0,58	0,60	0,58	0,62	0,53	0,58	0,62	0,51	1	6	7	4,5	8,5	3	4,5	8,5	2
8	1-NN	N3	manhattan	0,57	0,60	0,57	0,58	0,60	0,53	0,58	0,60	0,53	3,5	9	3,5	5,5	7,5	1	5,5	7,5	2
9	1-NN	N4	euclidean	0,46	0,58	0,55	0,53	0,64	0,55	0,53	0,64	0,51	1	7	6	3,5	8,5	5	3,5	8,5	2
10	1-NN	N4	manhattan	0,53	0,60	0,59	0,58	0,60	0,53	0,58	0,60	0,53	1	9	6	4,5	7,5	2	4,5	7,5	3
11	B-NN	N0	euclidean	0,47	0,46	0,47	0,47	0,62	0,46	0,47	0,62	0,51	4,5	2	4,5	4,5	8,5	1	4,5	8,5	7
12	B-NN	N0	manhattan	0,44	0,44	0,53	0,49	0,60	0,46	0,49	0,60	0,51	2	1	7	4,5	8,5	3	4,5	8,5	6
13	B-NN	N1	euclidean	0,56	0,58	0,58	0,58	0,57	0,47	0,58	0,57	0,54	3	7	9	7	4,5	1	7	4,5	2
14	B-NN	N1	manhattan	0,58	0,62	0,60	0,58	0,57	0,46	0,58	0,57	0,49	7	9	8	5,5	3,5	1	5,5	3,5	2
15	B-NN	N2	euclidean	0,60	0,58	0,58	0,60	0,57	0,48	0,60	0,57	0,59	7	4	5	8,5	2,5	1	8,5	2,5	6
16	B-NN	N2	manhattan	0,62	0,62	0,60	0,58	0,57	0,46	0,58	0,57	0,54	8	9	7	5,5	3,5	1	5,5	3,5	2
17	B-NN	N3	euclidean	0,57	0,58	0,60	0,55	0,62	0,48	0,55	0,62	0,59	4	5	7	2,5	8,5	1	2,5	8,5	6
18	B-NN	N3	manhattan	0,53	0,60	0,57	0,63	0,62	0,48	0,63	0,62	0,50	3	5	4	8,5	6,5	1	8,5	6,5	2
19	B-NN	N4	euclidean	0,62	0,58	0,60	0,66	0,62	0,48	0,66	0,62	0,54	7	3	4	8,5	5,5	1	8,5	5,5	2
20	B-NN	N4	manhattan	0,58	0,60	0,59	0,61	0,62	0,48	0,61	0,62	0,50	3	5	4	6,5	8,5	1	6,5	8,5	2
												Rerata	3,70	5,45	5,85	5,78	6,55	2,25	5,78	6,55	3,10
												Simp. Baku	2,18	2,72	2,06	1,92	2,21	1,59	1,92	2,21	2,15

Lampiran 4 Blok Pengujian Pengaruh Fitur pada Rerata Harmonik AUC dan *F-Measure* (Klasifikasi 4 Kelas)

Blok	Blok Penelitian			Pengaruh Skema Fitur									Peringkat								
	SNN	SND	Jarak	F1	F2	F3	F4	F5	F6	F7	F8	F9	F1	F2	F3	F4	F5	F6	F7	F8	F9
1	1-NN	N0	euclidean	0,43	0,33	0,43	0,39	0,40	0,37	0,41	0,42	0,39	8,5	1	8,5	3,5	5	2	6	7	3,5
2	1-NN	N0	manhattan	0,45	0,35	0,45	0,42	0,40	0,34	0,48	0,43	0,42	7,5	2	7,5	4,5	3	1	9	6	4,5
3	1-NN	N1	euclidean	0,41	0,54	0,47	0,41	0,40	0,36	0,43	0,42	0,41	3	9	8	4,5	2	1	7	6	4,5
4	1-NN	N1	manhattan	0,41	0,52	0,46	0,45	0,40	0,32	0,49	0,43	0,45	3	9	7	5,5	2	1	8	4	5,5
5	1-NN	N2	euclidean	0,41	0,54	0,47	0,41	0,40	0,36	0,43	0,42	0,41	3	9	8	4,5	2	1	7	6	4,5
6	1-NN	N2	manhattan	0,41	0,52	0,46	0,45	0,40	0,32	0,49	0,43	0,45	3	9	7	5,5	2	1	8	4	5,5
7	1-NN	N3	euclidean	0,46	0,51	0,47	0,42	0,40	0,36	0,43	0,42	0,42	7	9	8	4,5	2	1	6	3	4,5
8	1-NN	N3	manhattan	0,42	0,49	0,47	0,48	0,40	0,34	0,48	0,42	0,48	4	9	5	7,5	2	1	6	3	7,5
9	1-NN	N4	euclidean	0,45	0,53	0,47	0,45	0,41	0,36	0,47	0,45	0,45	6	9	7	4,5	2	1	8	3	4,5
10	1-NN	N4	manhattan	0,43	0,53	0,47	0,49	0,41	0,34	0,50	0,43	0,49	4	9	5	6,5	2	1	8	3	6,5
11	B-NN	N0	euclidean	0,43	0,40	0,43	0,43	0,41	0,42	0,41	0,42	0,43	8,5	1	8,5	6,5	3	4	2	5	6,5
12	B-NN	N0	manhattan	0,45	0,40	0,45	0,42	0,41	0,41	0,46	0,41	0,42	7,5	1	7,5	5,5	3	4	9	2	5,5
13	B-NN	N1	euclidean	0,42	0,54	0,42	0,50	0,41	0,41	0,50	0,48	0,50	3	9	4	7	2	1	7	5	7
14	B-NN	N1	manhattan	0,44	0,54	0,40	0,46	0,41	0,40	0,45	0,47	0,48	4	9	1	6	3	2	5	7	8
15	B-NN	N2	euclidean	0,42	0,54	0,42	0,50	0,41	0,41	0,50	0,48	0,48	3	9	4	7,5	2	1	7,5	5	6
16	B-NN	N2	manhattan	0,44	0,52	0,40	0,45	0,41	0,38	0,45	0,41	0,46	5	9	2	6	4	1	7	3	8
17	B-NN	N3	euclidean	0,42	0,51	0,43	0,48	0,41	0,41	0,48	0,42	0,48	4	9	5	7,5	2	1	6	3	7,5
18	B-NN	N3	manhattan	0,45	0,53	0,46	0,48	0,41	0,39	0,46	0,47	0,46	3	9	6	8	2	1	4	7	5
19	B-NN	N4	euclidean	0,40	0,53	0,43	0,47	0,41	0,41	0,46	0,41	0,47	1	9	5	7,5	4	3	6	2	7,5
20	B-NN	N4	manhattan	0,46	0,53	0,41	0,46	0,41	0,39	0,44	0,47	0,46	5	9	2	6,5	3	1	4	8	6,5
												Rerata	4,65	7,45	5,80	5,95	2,60	1,50	6,53	4,60	5,93
												Simp. Baku	2,15	3,19	2,30	1,32	0,88	1,00	1,77	1,85	1,37

Lampiran 5 Blok Pengujian Pengaruh Skema Nilai-K pada Rerata Harmonik  
AUC dan *F-Measure* (Klasifikasi 2 Kelas)

Blok	Blok Penelitian (Kombinasi Skema Lain)			Pengaruh Skema Nilai-K (RH)		Peringkat	
	Normalisasi	Fitur	Jarak	1-NN	B-NN	1-NN	B-NN
1	N0	F1	euclidean	0,4887	0,4659	2	1
2	N0	F1	manhattan	0,5333	0,4427	2	1
3	N0	F2	euclidean	0,4629	0,4629	1,5	1,5
4	N0	F2	manhattan	0,4419	0,4419	1,5	1,5
5	N0	F3	euclidean	0,4887	0,4659	2	1
6	N0	F3	manhattan	0,5333	0,5333	1,5	1,5
7	N0	F4	euclidean	0,5720	0,4659	2	1
8	N0	F4	manhattan	0,5510	0,4874	2	1
9	N0	F5	euclidean	0,5984	0,6173	1	2
10	N0	F5	manhattan	0,6223	0,5961	2	1
11	N0	F6	euclidean	0,5536	0,4608	2	1
12	N0	F6	manhattan	0,5778	0,4608	2	1
13	N0	F7	euclidean	0,5720	0,4659	2	1
14	N0	F7	manhattan	0,5510	0,4874	2	1
15	N0	F8	euclidean	0,5984	0,6173	1	2
16	N0	F8	manhattan	0,6223	0,5961	2	1
17	N0	F9	euclidean	0,5554	0,5059	2	1
18	N0	F9	manhattan	0,6217	0,5088	2	1
19	N1	F1	euclidean	0,5098	0,5557	1	2
20	N1	F1	manhattan	0,5942	0,5779	2	1
21	N1	F2	euclidean	0,5778	0,5778	1,5	1,5
22	N1	F2	manhattan	0,6223	0,6223	1,5	1,5
23	N1	F3	euclidean	0,6433	0,5779	2	1
24	N1	F3	manhattan	0,5969	0,5969	1,5	1,5
25	N1	F4	euclidean	0,5768	0,5778	1	2
26	N1	F4	manhattan	0,6224	0,5768	2	1
27	N1	F5	euclidean	0,5779	0,5749	2	1
28	N1	F5	manhattan	0,5557	0,5749	1	2
29	N1	F6	euclidean	0,5768	0,4659	2	1
30	N1	F6	manhattan	0,5551	0,4608	2	1
31	N1	F7	euclidean	0,5768	0,5778	1	2
32	N1	F7	manhattan	0,6224	0,5768	2	1
33	N1	F8	euclidean	0,5779	0,5749	2	1
34	N1	F8	manhattan	0,5557	0,5749	1	2
35	N1	F9	euclidean	0,4668	0,5359	1	2
36	N1	F9	manhattan	0,5333	0,4890	2	1
37	N2	F1	euclidean	0,5098	0,5999	1	2
38	N2	F1	manhattan	0,5942	0,6217	1	2
39	N2	F2	euclidean	0,5778	0,5778	1,5	1,5
40	N2	F2	manhattan	0,6223	0,6223	1,5	1,5
41	N2	F3	euclidean	0,6433	0,5779	2	1

Lampiran 5 Blok Pengujian Pengaruh Skema Nilai-K pada Rerata Harmonik  
AUC dan *F-Measure* (Klasifikasi 2 Kelas) - Lanjutan

Blok	Blok Penelitian (Kombinasi Skema Lain)			Pengaruh Skema Nilai-K (RH)		Peringkat	
	Normalisasi	Fitur	Jarak	1-NN	B-NN	1-NN	B-NN
42	N2	F3	manhattan	0,5969	0,5969	1,5	1,5
43	N2	F4	euclidean	0,5768	0,6002	1	2
44	N2	F4	manhattan	0,6224	0,5768	2	1
45	N2	F5	euclidean	0,5779	0,5749	2	1
46	N2	F5	manhattan	0,5557	0,5749	1	2
47	N2	F6	euclidean	0,5768	0,4850	2	1
48	N2	F6	manhattan	0,5551	0,4608	2	1
49	N2	F7	euclidean	0,5768	0,6002	1	2
50	N2	F7	manhattan	0,6224	0,5768	2	1
51	N2	F8	euclidean	0,5779	0,5749	2	1
52	N2	F8	manhattan	0,5557	0,5749	1	2
53	N2	F9	euclidean	0,4668	0,5929	1	2
54	N2	F9	manhattan	0,5333	0,5424	1	2
55	N3	F1	euclidean	0,5076	0,5732	1	2
56	N3	F1	manhattan	0,5732	0,5311	2	1
57	N3	F2	euclidean	0,5779	0,5779	1,5	1,5
58	N3	F2	manhattan	0,6002	0,6002	1,5	1,5
59	N3	F3	euclidean	0,5988	0,5988	1,5	1,5
60	N3	F3	manhattan	0,5732	0,5732	1,5	1,5
61	N3	F4	euclidean	0,5757	0,5543	2	1
62	N3	F4	manhattan	0,5757	0,6315	1	2
63	N3	F5	euclidean	0,6223	0,6223	1,5	1,5
64	N3	F5	manhattan	0,5997	0,6173	1	2
65	N3	F6	euclidean	0,5321	0,4850	2	1
66	N3	F6	manhattan	0,5321	0,4850	2	1
67	N3	F7	euclidean	0,5757	0,5543	2	1
68	N3	F7	manhattan	0,5757	0,6315	1	2
69	N3	F8	euclidean	0,6223	0,6223	1,5	1,5
70	N3	F8	manhattan	0,5997	0,6173	1	2
71	N3	F9	euclidean	0,5112	0,5885	1	2
72	N3	F9	manhattan	0,5335	0,5018	2	1
73	N4	F1	euclidean	0,4593	0,6203	1	2
74	N4	F1	manhattan	0,5311	0,5772	1	2
75	N4	F2	euclidean	0,5779	0,5779	1,5	1,5
76	N4	F2	manhattan	0,6002	0,6002	1,5	1,5
77	N4	F3	euclidean	0,5543	0,5988	1	2
78	N4	F3	manhattan	0,5942	0,5942	1,5	1,5
79	N4	F4	euclidean	0,5311	0,6601	1	2
80	N4	F4	manhattan	0,5757	0,6150	1	2
81	N4	F5	euclidean	0,6443	0,6173	2	1
82	N4	F5	manhattan	0,5997	0,6173	1	2
83	N4	F6	euclidean	0,5536	0,4850	2	1



Lampiran 5 Blok Pengujian Pengaruh Skema Nilai-K pada Rerata Harmonik AUC dan *F-Measure* (Klasifikasi 2 Kelas) - Lanjutan

Blok	Blok Penelitian (Kombinasi Skema Lain)			Pengaruh Skema Nilai-K (RH)		Peringkat	
	Normalisasi	Fitur	Jarak	1-NN	B-NN	1-NN	B-NN
84	N4	F6	manhattan	0,5321	0,4850	2	1
85	N4	F7	euclidean	0,5311	0,6601	1	2
86	N4	F7	manhattan	0,5757	0,6150	1	2
87	N4	F8	euclidean	0,6443	0,6173	2	1
88	N4	F8	manhattan	0,5997	0,6173	1	2
89	N4	F9	euclidean	0,5112	0,5359	1	2
90	N4	F9	manhattan	0,5335	0,5018	2	1
Rerata						1,54	1,46
Simp. Baku						0,45	0,45

Lampiran 6 Blok Pengujian Pengaruh Skema Nilai-K pada Rerata Harmonik AUC dan *F-Measure* (Klasifikasi 4 Kelas)

Blok	Blok Penelitian (Kombinasi Skema Lain)			Pengaruh Skema Nilai-K (RH)		Peringkat	
	Normalisasi	Fitur	Jarak	1-NN	B-NN	1-NN	B-NN
1	N0	F1	euclidean	0,4260	0,4260	1,5	1,5
2	N0	F1	manhattan	0,4502	0,4502	1,5	1,5
3	N0	F2	euclidean	0,3284	0,4000	1	2
4	N0	F2	manhattan	0,3519	0,4000	1	2
5	N0	F3	euclidean	0,4260	0,4260	1,5	1,5
6	N0	F3	manhattan	0,4502	0,4502	1,5	1,5
7	N0	F4	euclidean	0,3902	0,4256	1	2
8	N0	F4	manhattan	0,4246	0,4246	1,5	1,5
9	N0	F5	euclidean	0,4001	0,4140	1	2
10	N0	F5	manhattan	0,4001	0,4140	1	2
11	N0	F6	euclidean	0,3696	0,4172	1	2
12	N0	F6	manhattan	0,3386	0,4142	1	2
13	N0	F7	euclidean	0,4067	0,4067	1,5	1,5
14	N0	F7	manhattan	0,4772	0,4607	2	1
15	N0	F8	euclidean	0,4204	0,4204	1,5	1,5
16	N0	F8	manhattan	0,4257	0,4120	2	1
17	N0	F9	euclidean	0,3902	0,4256	1	2
18	N0	F9	manhattan	0,4246	0,4246	1,5	1,5
19	N1	F1	euclidean	0,4129	0,4186	1	2
20	N1	F1	manhattan	0,4096	0,4375	1	2
21	N1	F2	euclidean	0,5352	0,5352	1,5	1,5
22	N1	F2	manhattan	0,5212	0,5378	1	2

Lampiran 6 Blok Pengujian Pengaruh Skema Nilai-K pada Rerata Harmonik AUC dan *F-Measure* (Klasifikasi 4 Kelas) - Lanjutan

Blok	Blok Penelitian (Kombinasi Skema Lain)			Pengaruh Skema Nilai-K (RH)		Peringkat	
	Normalisasi	Fitur	Jarak	1-NN	B-NN	1-NN	B-NN
23	N1	F3	euclidean	0,4685	0,4201	2	1
24	N1	F3	manhattan	0,4646	0,4000	2	1
25	N1	F4	euclidean	0,4147	0,5001	1	2
26	N1	F4	manhattan	0,4505	0,4587	1	2
27	N1	F5	euclidean	0,4001	0,4140	1	2
28	N1	F5	manhattan	0,4001	0,4140	1	2
29	N1	F6	euclidean	0,3584	0,4083	1	2
30	N1	F6	manhattan	0,3247	0,4017	1	2
31	N1	F7	euclidean	0,4314	0,5001	1	2
32	N1	F7	manhattan	0,4949	0,4487	2	1
33	N1	F8	euclidean	0,4204	0,4832	1	2
34	N1	F8	manhattan	0,4282	0,4734	1	2
35	N1	F9	euclidean	0,4147	0,5001	1	2
36	N1	F9	manhattan	0,4505	0,4756	1	2
37	N2	F1	euclidean	0,4129	0,4186	1	2
38	N2	F1	manhattan	0,4096	0,4375	1	2
39	N2	F2	euclidean	0,5352	0,5352	1,5	1,5
40	N2	F2	manhattan	0,5212	0,5212	1,5	1,5
41	N2	F3	euclidean	0,4685	0,4201	2	1
42	N2	F3	manhattan	0,4646	0,4000	2	1
43	N2	F4	euclidean	0,4147	0,5001	1	2
44	N2	F4	manhattan	0,4505	0,4464	2	1
45	N2	F5	euclidean	0,4001	0,4140	1	2
46	N2	F5	manhattan	0,4001	0,4140	1	2
47	N2	F6	euclidean	0,3584	0,4112	1	2
48	N2	F6	manhattan	0,3247	0,3819	1	2
49	N2	F7	euclidean	0,4314	0,5001	1	2
50	N2	F7	manhattan	0,4949	0,4487	2	1
51	N2	F8	euclidean	0,4204	0,4832	1	2
52	N2	F8	manhattan	0,4282	0,4079	2	1
53	N2	F9	euclidean	0,4147	0,4842	1	2
54	N2	F9	manhattan	0,4505	0,4587	1	2
55	N3	F1	euclidean	0,4591	0,4213	2	1
56	N3	F1	manhattan	0,4242	0,4491	1	2
57	N3	F2	euclidean	0,5129	0,5129	1,5	1,5
58	N3	F2	manhattan	0,4899	0,5346	1	2
59	N3	F3	euclidean	0,4722	0,4293	2	1
60	N3	F3	manhattan	0,4660	0,4643	2	1
61	N3	F4	euclidean	0,4214	0,4768	1	2
62	N3	F4	manhattan	0,4781	0,4757	2	1
63	N3	F5	euclidean	0,4001	0,4140	1	2
64	N3	F5	manhattan	0,4001	0,4140	1	2
65	N3	F6	euclidean	0,3565	0,4106	1	2

Lampiran 6 Blok Pengujian Pengaruh Skema Nilai-K pada Rerata Harmonik AUC dan *F-Measure* (Klasifikasi 4 Kelas) - Lanjutan

Blok	Blok Penelitian (Kombinasi Skema Lain)			Pengaruh Skema Nilai-K (RH)		Peringkat	
	Normalisasi	Fitur	Jarak	1-NN	B-NN	1-NN	B-NN
66	N3	F6	manhattan	0,3425	0,3889	1	2
67	N3	F7	euclidean	0,4344	0,4757	1	2
68	N3	F7	manhattan	0,4761	0,4606	2	1
69	N3	F8	euclidean	0,4204	0,4204	1,5	1,5
70	N3	F8	manhattan	0,4180	0,4734	1	2
71	N3	F9	euclidean	0,4214	0,4768	1	2
72	N3	F9	manhattan	0,4781	0,4641	2	1
73	N4	F1	euclidean	0,4492	0,4026	2	1
74	N4	F1	manhattan	0,4293	0,4551	1	2
75	N4	F2	euclidean	0,5296	0,5296	1,5	1,5
76	N4	F2	manhattan	0,5341	0,5309	2	1
77	N4	F3	euclidean	0,4653	0,4264	2	1
78	N4	F3	manhattan	0,4738	0,4086	2	1
79	N4	F4	euclidean	0,4466	0,4694	1	2
80	N4	F4	manhattan	0,4865	0,4563	2	1
81	N4	F5	euclidean	0,4066	0,4140	1	2
82	N4	F5	manhattan	0,4066	0,4140	1	2
83	N4	F6	euclidean	0,3613	0,4106	1	2
84	N4	F6	manhattan	0,3396	0,3889	1	2
85	N4	F7	euclidean	0,4665	0,4591	2	1
86	N4	F7	manhattan	0,4995	0,4444	2	1
87	N4	F8	euclidean	0,4459	0,4079	2	1
88	N4	F8	manhattan	0,4252	0,4734	1	2
89	N4	F9	euclidean	0,4466	0,4694	1	2
90	N4	F9	manhattan	0,4865	0,4563	2	1
Rerata				0,43	0,45	1,36	1,64
Simp. Baku				0,05	0,04	0,44	0,44

Lampiran 7 Blok Pengujian Pengaruh Skema Jarak pada Rerata Harmonik AUC dan *F-Measure* (Klasifikasi 2 Kelas)

Blok	Blok Penelitian (Kombinasi Skema Lain)			Pengaruh Skema Jarak (RH)		Peringkat	
	SNN	SND	Fitur	Euclidean	Manhattan	Euclidean	Manhattan
1	1-NN	N0	F1	0,4887	0,5333	1	2
2	1-NN	N0	F2	0,4629	0,4419	2	1
3	1-NN	N0	F3	0,4887	0,5333	1	2
4	1-NN	N0	F4	0,5720	0,5510	2	1
5	1-NN	N0	F5	0,5984	0,6223	1	2

Lampiran 7 Blok Pengujian Pengaruh Skema Jarak pada Rerata Harmonik AUC dan *F-Measure* (Klasifikasi 2 Kelas) - Lanjutan

Blok	Blok Penelitian (Kombinasi Skema Lain)			Pengaruh Skema Jarak (RH)		Peringkat	
	SNN	SND	Fitur	Euclidean	Manhattan	Euclidean	Manhattan
6	1-NN	N0	F6	0,5536	0,5778	1	2
7	1-NN	N0	F7	0,5720	0,5510	2	1
8	1-NN	N0	F8	0,5984	0,6223	1	2
9	1-NN	N0	F9	0,5554	0,6217	1	2
10	1-NN	N1	F1	0,5098	0,5942	1	2
11	1-NN	N1	F2	0,5778	0,6223	1	2
12	1-NN	N1	F3	0,6433	0,5969	2	1
13	1-NN	N1	F4	0,5768	0,6224	1	2
14	1-NN	N1	F5	0,5779	0,5557	2	1
15	1-NN	N1	F6	0,5768	0,5551	2	1
16	1-NN	N1	F7	0,5768	0,6224	1	2
17	1-NN	N1	F8	0,5779	0,5557	2	1
18	1-NN	N1	F9	0,4668	0,5333	1	2
19	1-NN	N2	F1	0,5098	0,5942	1	2
20	1-NN	N2	F2	0,5778	0,6223	1	2
21	1-NN	N2	F3	0,6433	0,5969	2	1
22	1-NN	N2	F4	0,5768	0,6224	1	2
23	1-NN	N2	F5	0,5779	0,5557	2	1
24	1-NN	N2	F6	0,5768	0,5551	2	1
25	1-NN	N2	F7	0,5768	0,6224	1	2
26	1-NN	N2	F8	0,5779	0,5557	2	1
27	1-NN	N2	F9	0,4668	0,5333	1	2
28	1-NN	N3	F1	0,5076	0,5732	1	2
29	1-NN	N3	F2	0,5779	0,6002	1	2
30	1-NN	N3	F3	0,5988	0,5732	2	1
31	1-NN	N3	F4	0,5757	0,5757	1,5	1,5
32	1-NN	N3	F5	0,6223	0,5997	2	1
33	1-NN	N3	F6	0,5321	0,5321	1,5	1,5
34	1-NN	N3	F7	0,5757	0,5757	1,5	1,5
35	1-NN	N3	F8	0,6223	0,5997	2	1
36	1-NN	N3	F9	0,5112	0,5335	1	2
37	1-NN	N4	F1	0,4593	0,5311	1	2
38	1-NN	N4	F2	0,5779	0,6002	1	2
39	1-NN	N4	F3	0,5543	0,5942	1	2
40	1-NN	N4	F4	0,5311	0,5757	1	2
41	1-NN	N4	F5	0,6443	0,5997	2	1
42	1-NN	N4	F6	0,5536	0,5321	2	1
43	1-NN	N4	F7	0,5311	0,5757	1	2
44	1-NN	N4	F8	0,6443	0,5997	2	1

Lampiran 7 Blok Pengujian Pengaruh Skema Jarak pada Rerata Harmonik AUC dan *F-Measure* (Klasifikasi 2 Kelas) - Lanjutan

Blok	Blok Penelitian (Kombinasi Skema Lain)			Pengaruh Skema Jarak (RH)		Peringkat	
	SNN	SND	Fitur	Euclidean	Manhattan	Euclidean	Manhattan
45	1-NN	N4	F9	0,5112	0,5335	1	2
46	B-NN	N0	F1	0,4659	0,4427	2	1
47	B-NN	N0	F2	0,4629	0,4419	2	1
48	B-NN	N0	F3	0,4659	0,5333	1	2
49	B-NN	N0	F4	0,4659	0,4874	1	2
50	B-NN	N0	F5	0,6173	0,5961	2	1
51	B-NN	N0	F6	0,4608	0,4608	1,5	1,5
52	B-NN	N0	F7	0,4659	0,4874	1	2
53	B-NN	N0	F8	0,6173	0,5961	2	1
54	B-NN	N0	F9	0,5059	0,5088	1	2
55	B-NN	N1	F1	0,5557	0,5779	1	2
56	B-NN	N1	F2	0,5778	0,6223	1	2
57	B-NN	N1	F3	0,5779	0,5969	1	2
58	B-NN	N1	F4	0,5778	0,5768	2	1
59	B-NN	N1	F5	0,5749	0,5749	1,5	1,5
60	B-NN	N1	F6	0,4659	0,4608	2	1
61	B-NN	N1	F7	0,5778	0,5768	2	1
62	B-NN	N1	F8	0,5749	0,5749	1,5	1,5
63	B-NN	N1	F9	0,5359	0,4890	2	1
64	B-NN	N2	F1	0,5999	0,6217	1	2
65	B-NN	N2	F2	0,5778	0,6223	1	2
66	B-NN	N2	F3	0,5779	0,5969	1	2
67	B-NN	N2	F4	0,6002	0,5768	2	1
68	B-NN	N2	F5	0,5749	0,5749	1,5	1,5
69	B-NN	N2	F6	0,4850	0,4608	2	1
70	B-NN	N2	F7	0,6002	0,5768	2	1
71	B-NN	N2	F8	0,5749	0,5749	1,5	1,5
72	B-NN	N2	F9	0,5929	0,5424	2	1
73	B-NN	N3	F1	0,5732	0,5311	2	1
74	B-NN	N3	F2	0,5779	0,6002	1	2
75	B-NN	N3	F3	0,5988	0,5732	2	1
76	B-NN	N3	F4	0,5543	0,6315	1	2
77	B-NN	N3	F5	0,6223	0,6173	2	1
78	B-NN	N3	F6	0,4850	0,4850	1,5	1,5
79	B-NN	N3	F7	0,5543	0,6315	1	2
80	B-NN	N3	F8	0,6223	0,6173	2	1
81	B-NN	N3	F9	0,5885	0,5018	2	1
82	B-NN	N4	F1	0,6203	0,5772	2	1
83	B-NN	N4	F2	0,5779	0,6002	1	2

Lampiran 7 Blok Pengujian Pengaruh Skema Jarak pada Rerata Harmonik AUC dan *F-Measure* (Klasifikasi 2 Kelas) - Lanjutan

Blok	Blok Penelitian (Kombinasi Skema Lain)			Pengaruh Skema Jarak (RH)		Peringkat	
	SNN	SND	Fitur	Euclidean	Manhattan	Euclidean	Manhattan
84	B-NN	N4	F3	0,5988	0,5942	2	1
85	B-NN	N4	F4	0,6601	0,6150	2	1
86	B-NN	N4	F5	0,6173	0,6173	1,5	1,5
87	B-NN	N4	F6	0,4850	0,4850	1,5	1,5
88	B-NN	N4	F7	0,6601	0,6150	2	1
89	B-NN	N4	F8	0,6173	0,6173	1,5	1,5
90	B-NN	N4	F9	0,5359	0,5018	2	1
Rerata						1,50	1,50
Simp. Baku						0,47	0,47

Lampiran 8 Blok Pengujian Pengaruh Skema Jarak pada Rerata Harmonik AUC dan *F-Measure* (Klasifikasi 4 Kelas)

Blok	Blok Penelitian (Kombinasi Skema Lain)			Pengaruh Skema Jarak (RH)		Peringkat	
	SNN	SND	Fitur	Euclidean	Manhattan	Euclidean	Manhattan
1	1-NN	N0	F1	0,4260	0,4502	1	2
2	1-NN	N0	F2	0,3284	0,3519	1	2
3	1-NN	N0	F3	0,4260	0,4502	1	2
4	1-NN	N0	F4	0,3902	0,4246	1	2
5	1-NN	N0	F5	0,4001	0,4001	1,5	1,5
6	1-NN	N0	F6	0,3696	0,3386	2	1
7	1-NN	N0	F7	0,4067	0,4772	1	2
8	1-NN	N0	F8	0,4204	0,4257	1	2
9	1-NN	N0	F9	0,3902	0,4246	1	2
10	1-NN	N1	F1	0,4129	0,4096	2	1
11	1-NN	N1	F2	0,5352	0,5212	2	1
12	1-NN	N1	F3	0,4685	0,4646	2	1
13	1-NN	N1	F4	0,4147	0,4505	1	2
14	1-NN	N1	F5	0,4001	0,4001	1,5	1,5
15	1-NN	N1	F6	0,3584	0,3247	2	1
16	1-NN	N1	F7	0,4314	0,4949	1	2
17	1-NN	N1	F8	0,4204	0,4282	1	2
18	1-NN	N1	F9	0,4147	0,4505	1	2
19	1-NN	N2	F1	0,4129	0,4096	2	1
20	1-NN	N2	F2	0,5352	0,5212	2	1
21	1-NN	N2	F3	0,4685	0,4646	2	1
22	1-NN	N2	F4	0,4147	0,4505	1	2
23	1-NN	N2	F5	0,4001	0,4001	1,5	1,5

Lampiran 8 Blok Pengujian Pengaruh Skema Jarak pada Rerata Harmonik AUC dan *F-Measure* (Klasifikasi 4 Kelas) - Lanjutan

Blok	Blok Penelitian (Kombinasi Skema Lain)			Pengaruh Skema Jarak (RH)		Peringkat	
	SNN	SND	Fitur	Euclidean	Manhattan	Euclidean	Manhattan
24	1-NN	N2	F6	0,3584	0,3247	2	1
25	1-NN	N2	F7	0,4314	0,4949	1	2
26	1-NN	N2	F8	0,4204	0,4282	1	2
27	1-NN	N2	F9	0,4147	0,4505	1	2
28	1-NN	N3	F1	0,4591	0,4242	2	1
29	1-NN	N3	F2	0,5129	0,4899	2	1
30	1-NN	N3	F3	0,4722	0,4660	2	1
31	1-NN	N3	F4	0,4214	0,4781	1	2
32	1-NN	N3	F5	0,4001	0,4001	1,5	1,5
33	1-NN	N3	F6	0,3565	0,3425	2	1
34	1-NN	N3	F7	0,4344	0,4761	1	2
35	1-NN	N3	F8	0,4204	0,4180	2	1
36	1-NN	N3	F9	0,4214	0,4781	1	2
37	1-NN	N4	F1	0,4492	0,4293	2	1
38	1-NN	N4	F2	0,5296	0,5341	1	2
39	1-NN	N4	F3	0,4653	0,4738	1	2
40	1-NN	N4	F4	0,4466	0,4865	1	2
41	1-NN	N4	F5	0,4066	0,4066	1,5	1,5
42	1-NN	N4	F6	0,3613	0,3396	2	1
43	1-NN	N4	F7	0,4665	0,4995	1	2
44	1-NN	N4	F8	0,4459	0,4252	2	1
45	1-NN	N4	F9	0,4466	0,4865	1	2
46	B-NN	N0	F1	0,4260	0,4502	1	2
47	B-NN	N0	F2	0,4000	0,4000	1,5	1,5
48	B-NN	N0	F3	0,4260	0,4502	1	2
49	B-NN	N0	F4	0,4256	0,4246	2	1
50	B-NN	N0	F5	0,4140	0,4140	1,5	1,5
51	B-NN	N0	F6	0,4172	0,4142	2	1
52	B-NN	N0	F7	0,4067	0,4607	1	2
53	B-NN	N0	F8	0,4204	0,4120	2	1
54	B-NN	N0	F9	0,4256	0,4246	2	1
55	B-NN	N1	F1	0,4186	0,4375	1	2
56	B-NN	N1	F2	0,5352	0,5378	1	2
57	B-NN	N1	F3	0,4201	0,4000	2	1
58	B-NN	N1	F4	0,5001	0,4587	2	1
59	B-NN	N1	F5	0,4140	0,4140	1,5	1,5
60	B-NN	N1	F6	0,4083	0,4017	2	1
61	B-NN	N1	F7	0,5001	0,4487	2	1
62	B-NN	N1	F8	0,4832	0,4734	2	1
63	B-NN	N1	F9	0,5001	0,4756	2	1
64	B-NN	N2	F1	0,4186	0,4375	1	2
65	B-NN	N2	F2	0,5352	0,5212	2	1

Lampiran 8 Blok Pengujian Pengaruh Skema Jarak pada Rerata Harmonik AUC dan *F-Measure* (Klasifikasi 4 Kelas) - Lanjutan

Blok	Blok Penelitian (Kombinasi Skema Lain)			Pengaruh Skema Jarak (RH)		Peringkat	
	SNN	SND	Fitur	Euclidean	Manhattan	Euclidean	Manhattan
66	B-NN	N2	F3	0,4201	0,4000	2	1
67	B-NN	N2	F4	0,5001	0,4464	2	1
68	B-NN	N2	F5	0,4140	0,4140	1,5	1,5
69	B-NN	N2	F6	0,4112	0,3819	2	1
70	B-NN	N2	F7	0,5001	0,4487	2	1
71	B-NN	N2	F8	0,4832	0,4079	2	1
72	B-NN	N2	F9	0,4842	0,4587	2	1
73	B-NN	N3	F1	0,4213	0,4491	1	2
74	B-NN	N3	F2	0,5129	0,5346	1	2
75	B-NN	N3	F3	0,4293	0,4643	1	2
76	B-NN	N3	F4	0,4768	0,4757	2	1
77	B-NN	N3	F5	0,4140	0,4140	1,5	1,5
78	B-NN	N3	F6	0,4106	0,3889	2	1
79	B-NN	N3	F7	0,4757	0,4606	2	1
80	B-NN	N3	F8	0,4204	0,4734	1	2
81	B-NN	N3	F9	0,4768	0,4641	2	1
82	B-NN	N4	F1	0,4026	0,4551	1	2
83	B-NN	N4	F2	0,5296	0,5309	1	2
84	B-NN	N4	F3	0,4264	0,4086	2	1
85	B-NN	N4	F4	0,4694	0,4563	2	1
86	B-NN	N4	F5	0,4140	0,4140	1,5	1,5
87	B-NN	N4	F6	0,4106	0,3889	2	1
88	B-NN	N4	F7	0,4591	0,4444	2	1
89	B-NN	N4	F8	0,4079	0,4734	1	2
90	B-NN	N4	F9	0,4694	0,4563	2	1
					Rerata	1,54	1,46
					Simp. Baku	0,47	0,47



Lampiran 9 Blok Pengujian Pengaruh Skema Uji Lanjut pada Rerata Harmonik  
AUC dan *F-Measure* (Klasifikasi 2 Kelas)

Blok	Blok Penelitian (Kombinasi Skema Lain)			Pengaruh Skema Uji Lanjut (RH)		Peringkat	
	SNN	Normalisasi	Jarak	Tukey	WMW	Tukey	WMW
1	1-NN	N0	euclidean	0,5746	0,5753	1	2
2	1-NN	N0	manhattan	0,5837	0,5983	1	2
3	1-NN	N1	euclidean	0,5772	0,5405	2	1
4	1-NN	N1	manhattan	0,5777	0,5705	2	1
5	1-NN	N2	euclidean	0,5772	0,5405	2	1
6	1-NN	N2	manhattan	0,5777	0,5705	2	1
7	1-NN	N3	euclidean	0,5767	0,5697	2	1
8	1-NN	N3	manhattan	0,5692	0,5696	1	2
9	1-NN	N4	euclidean	0,5763	0,5622	2	1
10	1-NN	N4	manhattan	0,5692	0,5696	1	2
11	B-NN	N0	euclidean	0,5147	0,5297	1	2
12	B-NN	N0	manhattan	0,5148	0,5308	1	2
13	B-NN	N1	euclidean	0,5395	0,5629	1	2
14	B-NN	N1	manhattan	0,5375	0,5469	1	2
15	B-NN	N2	euclidean	0,5534	0,5893	1	2
16	B-NN	N2	manhattan	0,5375	0,5647	1	2
17	B-NN	N3	euclidean	0,5539	0,5884	1	2
18	B-NN	N3	manhattan	0,5779	0,5835	1	2
19	B-NN	N4	euclidean	0,5875	0,6044	1	2
20	B-NN	N4	manhattan	0,5724	0,5780	1	2
					Rerata	1,30	1,70
					Simp. Baku	0,47	0,47

Lampiran 10 Blok Pengujian Pengaruh Skema Uji Lanjut pada Rerata Harmonik  
AUC dan *F-Measure* (Klasifikasi 4 Kelas)

Blok	Blok Penelitian (Kombinasi Skema Lain)			Pengaruh Skema Uji Lanjut (RH)		Peringkat	
	SNN	Normalisasi	Jarak	Tukey	WMW	Tukey	WMW
1	1-NN	N0	euclidean	0,3866	0,4058	1	2
2	1-NN	N0	manhattan	0,3877	0,4425	1	2
3	1-NN	N1	euclidean	0,3910	0,4221	1	2
4	1-NN	N1	manhattan	0,3918	0,4579	1	2
5	1-NN	N2	euclidean	0,3910	0,4221	1	2
6	1-NN	N2	manhattan	0,3918	0,4579	1	2
7	1-NN	N3	euclidean	0,3927	0,4254	1	2
8	1-NN	N3	manhattan	0,4069	0,4574	1	2
9	1-NN	N4	euclidean	0,4048	0,4530	1	2
10	1-NN	N4	manhattan	0,4109	0,4704	1	2
11	B-NN	N0	euclidean	0,4189	0,4176	2	1
12	B-NN	N0	manhattan	0,4176	0,4324	1	2
13	B-NN	N1	euclidean	0,4408	0,4945	1	2
14	B-NN	N1	manhattan	0,4248	0,4659	1	2
15	B-NN	N2	euclidean	0,4418	0,4892	1	2
16	B-NN	N2	manhattan	0,4141	0,4384	1	2
17	B-NN	N3	euclidean	0,4338	0,4576	1	2
18	B-NN	N3	manhattan	0,4262	0,4661	1	2
19	B-NN	N4	euclidean	0,4313	0,4455	1	2
20	B-NN	N4	manhattan	0,4197	0,4580	1	2
					Rerata	1,05	1,95
					Simp. Baku	0,22	0,22

## BIOGRAFI PENULIS



**Ahmad Dwi Arianto**, Magister di Institut Teknologi Sepuluh Nopember, Fakultas Teknologi Elektro, Departemen Teknik Elektro, Bidang Keahlian Telematika Chief Information Officer (CIO), merupakan salah satu penerima Beasiswa Kerjasama Kementerian Komunikasi dan Informasi Republik Indonesia dengan Badan

Pemeriksa Keuangan Republik Indonesia. Lahir di Blora pada tanggal 10 Maret 1986. Anak pertama dari 4 bersaudara dari pasangan Sunarto dan Nining Setyowati, suami dari Anisy Syahidah dan ayah dari Annisa Istifiyanza Bihurin'in Arianto. Bertugas pada Badan Pemeriksa Keuangan Republik Indonesia.

*Alamat Email:* [interimtani13@gmail.com](mailto:interimtani13@gmail.com)

### **Riwayat Pendidikan :**

**SD** : SDN Pilang 1, Blora, Jawa Tengah  
**SMP** : Mts Islam Al-Mukmin, Surakarta, Jawa Tengah  
**SMU** : SMUN 1 Randublatung, Blora, Jawa Tengah  
**D4/S1** : STAN, Tangerang Selatan, Banten